WORKING PAPER NO. 38

The Formation and Evolution of Childhood Skill Acquisition: Evidence From India

January 2009

Christian Helmers

Manasa Patnam

Young Lives
An International Study of Childhood Poverty

WORKING PAPER NO. 38

The Formation and Evolution of Childhood Skill Acquisition: Evidence From India

January 2009

Christian Helmers
Manasa Patnam

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

The Formation and Evolution of Childhood Skill Acquisition:
Evidence From India

Christian Helmers and Manasa Patnam

First published by Young Lives in January 2009

ISBN: 978-1-904427-43-8

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 289966
E-mail: younglives@younglives.org.uk
Web: www.younglives.org.uk

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

Contents

Abstract

Building on recent advances in the literature and using a rich data set for two cohorts of children aged between one and twelve for Andhra Pradesh, India, we investigate the determinants of children's cognitive as well as non-cognitive skills. We find evidence of self-productivity for cognitive skills and cross-productivity effects from cognitive on non-cognitive skills. Moreover, we demonstrate that parental investment has contemporaneously powerful positive effects on skill levels for all age groups. Investigating other determinants of these skills, we find child health at age one to influence cognitive abilities at age five, whilst child health at age one is influenced by parental care already during pregnancy and earliest childhood. Understanding the determinants which account explicitly for the effects of a large number of child, caregiver and household characteristics provides insights with regard to possible policy interventions to improve the chances of children in poor environments of developing cognitive and non-cognitive skills crucial for success in many spheres of life.

Keywords: Children, cognitive skills, non-cognitive skills
JEL Classification: J13, O15

Acknowledgements

We benefitted from comments from Maria Laura De Tommaso and Pia Pinger, participants in a seminar at Cambridge, participants to the LACEA / LAMES 2008 annual meeting, the 4th Annual Conference on Economic Growth and Development at the Indian Statistical Institute, and the ESDS International Annual Conference 2008. This research was funded by Young Lives (www.younglives.org.uk), a long-term international research project investigating the changing nature of childhood poverty.

The Authors

Christian Helmers is a DPhil student at the Department of Economics at Oxford University. He is also a Research Affiliate at the Kiel Institute for the World Economy in Germany. Christian Helmers earned his first degree in Economics from the University of Lausanne. He also holds a MSc from Oxford University and a postgraduate degree from the Kiel Institute. He has worked as a consultant for ITC UNCTAD/WTO, German Development Cooperation (GTZ), and various UK institutions.

Manasa Patnam is a PhD student at the Department of Economics, University Of Cambridge. She has earned her first degree in Economics from the University of Delhi. She also holds an MSc from Oxford University.

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.

1 Introduction

The investigation of the determinants of a child's socio-economic success in later life has traditionally focused on schooling. A large body of literature in psychology and more recently also in economics, however, argues that the true determinants of a child's success in life are formed already during early childhood. Heckman et al. (2006) have shown that for this reason, school quality and resources devoted to students are only effective in as much as they remedy (to a limited extent) deficits in ability caused in earlier childhood. In explaining school and professional success much emphasis has been put on the analysis of the importance of cognitive skills (Heckman, 1995). The psychometric literature (e.g., Heckman and Rubinstein, 2001; Cunha and Heckman, 2008), however, has shown that non-cognitive abilities, such as motivation, perseverance, risk aversion, self-control etc., also play an important role in setting the course for a successful later life beyond childhood.

The early economic analysis of child development (Becker and Tomes, 1986) assumed that childhood consisted of a single homogeneous phase. Psychological research (e.g., Thompson and Nelson, 2001) demonstrated that child development is comprised of multiple stages. A child develops cognitive and non-cognitive skills through different stages during childhood. She acquires cognitive skills very early in life, beginning already in the womb. The development of cognitive skills, for example IQ, is mostly completed by age eight to ten. While the development of non-cognitive skills also starts early in life and is equally affected by a child's environment, these skills remain malleable later in life (Carneiro, Cunha and Heckman, 2003). Importantly, these cognitive and non-cognitive skills interact and reinforce each other, characteristics termed *self-* and *cross-productivity* (Carneiro et al., 2003).

There are two main influences shaping a child's abilities during this multistage development process: a child's genetic endowment as well as inputs received from the outside world, including family and the wider environment children are born into. Many studies highlight the importance of genetic variance in determining child development. By examining behavioural patterns of adopted children vis-a-vis their siblings (Scarr and Richard, 1983; Teasdale and Owen, 1985), as well as those between a pair of monozygotic and dizygotic twins (Wilson, 1983), the research has focused on establishing that siblings with uncorrelated genetic structure do not resemble each other in any measured talents. Thus while family and home environment contexts may have common influences in fostering child development, these environments can be construed in different ways by children based on their genome structure. Nonetheless, these studies highlight how family and environment play a crucial role in child development over the child's early periods. Scarr and Weinberg (1983) find, from a sample of transracial families, that young siblings (black/interracial adoptees compared to their natural brothers/sisters) are intellectually similar, hence concluding that younger children are more influenced by differences within their family environment.[1] While the genetic endowment is given, a child's genetic expression is nevertheless influenced by the child's environment (Turkheimer et al., 2003). Emerging research by geneticists show that environmental factors may cause genes to express themselves differently. This effect, traditionally termed as 'epigentics', reveals that many behaviour patterns of genes can be altered by life cycle experiences (Hunter, 2008). Regarding inputs from outside, Carneiro and Heckman (2003) have shown that the principal source of influence are a child's parents. Hence, a child's abilities can be produced and modified through outside influences over the different stages of childhood and even to some extent in later adult life. This implies that inputs and environmental factors have different effects during these different development stages. In addition, inputs are cumulative in nature, a characteristic labeled dynamic complementarity (Carneiro et al., 2003). This, in turn, implies that early investments must be followed up by investments during later stages in life to render early investments more effective (Currie and Thomas, 1995). It also means that deficiencies, above all in cognitive skills, caused early in life are hard to remedy and the later remediation occurs, the less effective it is (O'Conner et al., 2004). Hence, while for some skills late remediation is nevertheless possible, it would be more efficient to avoid the emergence of these deficiencies early in life (Cunha and Heckman, 2006).

These findings about the development process of a child's abilities are important as Cunha et al. (2006) have found that children diverge very quickly in terms of their development of cognitive as well as non-cognitive skills according to the input received during the sensitive phases of early childhood. This has substantial and far-reaching implications for a child's later life given the large evidence on ability as a predictor for school success, wages, crime involvement and other spheres of life (Heckman, 1995; Heckman et al., 2006). We also know that returns to schooling and job training are lower for individuals with lower skill levels (Carneiro and Heckman, 2003), which justifies early intervention on efficiency grounds as high returns may be expected from remediation for disadvantaged children (Blau and Currie, 2006).

So far, the bulk of this research has been undertaken in industrialised countries. Little is known about the potentially enormous implications of these findings in the

[1]'A major reason for the greater resemblance [...] is that families have greater effects on their younger than older children' (Scarr, p. 11: 1992).

developing country context (Grantham-McGregor et al., 2007). In most developing countries school quality is despicably low, and while demand for schooling from households may be reflected as quite high in terms of school enrolments, it has been well documented that the achievement levels of children especially in rural schools still continues to lag (Gonzalez et al., 2004). In order to improve school performance and a child's chances of succeeding in later life, the literature discussed gives strong reason to shift focus from schooling towards other factors influencing a child's development of cognitive and non-cognitive traits during early childhood. Whether a child is ready for and achieves progress in school depends to a large extent on the child's cognitive attainment level, but also hinges to a non-negligible extent on her social and emotional development, a component mostly forgotten in the developing country context. In a recent study, Grantham-McGregor et al. (2007) find that both poverty and bad home environment condition result in child stuntedness hampering early child development. The authors go on to show how these early childhood disadvantages affect later period skill development: children who are stunted at early stages tend to perform poorly later on at school. Taking this analysis further, Walker et al. (2007) chart out channels that are most likely to affect early childhood development, leading to child stuntedness. Important factors, include what the authors term 'Psycho-social Risk Factors' (PRF) and childhood poverty. The authors describe PRF as a combination of parenting factors (child-learning opportunities, caregiver sensitivity), and contextual factors (maternal depression and exposure to violence). Hence, high PRF and childhood poverty can lead to childhood stuntedness which subsequently leads to adverse cognitive and non-cognitive skill acquisition outcomes. A previous analysis using the first round of data of the 'Young Lives' survey in India's state of Andhra Pradesh for eight-year-old children by Galab et. al. (2006) found that girls lagged in achievement compared to boys, children of uneducated parents are at a disadvantage, and more importantly, that there is a clear interplay between school and home. The authors assert that while it may be that child learning is determined at school, home environment remains crucial in fostering child achievement. They argue that children tend to perform better on literacy and numeracy tasks if parents place high value on schooling (therefore investing more in their child's education) and actively support/help their children at home.

Despite these recent advances in the literature, little is known as to how a child's cognitive and non-cognitive skills interact and how a child's environment, including family, affects cognitive as well as non-cognitive development in a developing country context. Moreover, while the economic literature on child development advanced by Heckman, Cunha and co-authors has emphasised the interplay between cognitive and non-cognitive skills and inputs stemming from a child's environment, the impact of social networks has been largely neglected so far. However, the literature on developing countries has shown the importance of social networks for poor households in coping with common adversities (Fafchamps and Gubert, 2006).

The objective of our analysis is therefore to investigate determinants of a child's development of cognitive and non-cognitive skills over various phases of childhood, paying particular attention to self- and cross-productivity effects within a developing country context. Self-productivity refers to any effect of past periods' cognitive/non-cognitive skills on the current period's cognitive/non-cognitive skills respectively, while cross-productivity refers to any effect of past periods' cognitive/non-cognitive skills on current period non-cognitive/cognitive skills. For this purpose, we estimate a Linear Structural Relations (LISREL) model which allows us to estimate latent cognitive and non-cognitive skill levels as well as parental investment and to link these variables to observed child, parental and household characteristics. Building on Cunha and Heckman (2007), we specifically examine the dynamics of both cognitive and non-cognitive skills as well as their interplay over time. In Section 2, we discuss in detail our methodology for the skill acquisition model. Section 3 describes the data used for our empirical investigation, while Section 4 lays out its estimation procedure. Section 5 presents our results and findings. Section 6 concludes.

2 Methodology

Our main interest lies in investigating the various determinants of skill formation with a particular focus on the effects of a child's past cognitive and non-cognitive skill levels in determining current levels of the same. This includes self-productivity as well as cross-productivity effects. In order to achieve this, we follow Cunha and Heckman (2008) in writing a child's skill level at age t as a function of the child's past level of skills, current parental investment, and other contemporaneous variables including child, caregiver, and household characteristics.

$$\theta_t^k = f(\theta_{t-1}^k, \theta_t^I, X_t) \tag{1}$$

where θ_t^k denotes a child's skill level of skill k for age t, with $k \in \{C, N\}$ and $t \in \{0, ..., T\}$. θ_t^I denotes parental investment at age t, and X_t denotes a vector of child, caregiver and household characteristics. Importantly, function (1) allows for self-productivity, i.e., θ_{t-1}^N and θ_{t-1}^C having an effect on θ_t^N and θ_t^C respectively, and cross-productivity, i.e., θ_{t-1}^N and θ_{t-1}^C having an effect on θ_t^C and θ_t^N respectively. In order to take account of the genetic endowment with which a child is born, we assume that each child is born with an individual initial set $\theta_o = (\theta_o^C, \theta_o^N)$. It is this initial set of skills which introduces - in principle unobserved - heterogeneity across children.[2] Todd and Wolpin (2006) propose to capture genetic endowment through a functional relationship such as

$$\theta_{it}^k = f(\theta_{it-1}^k, \theta_{it}^I, X_{it}, \mu_i) \tag{2}$$

where i denotes individuals. Todd and Wolpin (2006) propose to take first differences of a linear specification of (2) to eliminate μ_i where first differences of skill levels are interpreted as value-added. Purging a linear specification of (2) through first-differences relies on the fact that the *expression* of unobserved genetic endowment is constant over time - which appears a rather strong assumption. The problem with value-added specifications, as pointed out by Andrabi et. al. (2008), is that it fails to account for the dynamic nature of achievement gains, wherein individual level heterogeneity entering in each period interacts with the achievement level in the past period. Using data from Pakistan, the authors find that value-added specification perform worse than cross-sectional comparisons. But even if we could reasonably assume that the expression of child-specific genetic characteristics can be captured through a time-invariant fixed effect, estimating first differences implies we would not be able to estimate the effect of cumulative self-productivity and cross-productivity as we have data only for two time periods for each cohort of children, t and $t-1$. The approach proposed by Cunha, Heckman and Navarro (2005) and Cunha and Heckman (2008) allows us to take account of fixed effects through measurement equations estimating the latent variables. In particular, parental investment, which is a choice variable and would be contemporaneously correlated with the presence of μ_i, is treated as a latent variable and estimated in the measurement model. Our control variables discussed above can all be regarded as state variables and should therefore not be contemporaneously correlated with the fixed effect. Also some controls serve, to an extent, as a proxy for unobserved child heterogeneity with reference to the older cohort.[3] However, apart from children having unobserved genetic characteristics, there also exist problems with unobservability of

[2]Note that there is an important distinction between heterogeneity and uncertainty caused by unobserved genetic endowment. Heterogeneity refers to individuals making choices based on their knowledge about their specific type whereas uncertainty refers to a situation where the type is unknown (Browning and Carro, 2006). It appears, therefore, more appropriate to regard a child's genetic endowment for the one-year-old children in our data as causing uncertainty rather than heterogeneity.

[3]We use child anthropometries, specifically child height as one such control. In a recent study, Weedon et al. (2007) establish that height is a typical 'polygenic trait'. The study finds that common variants in the HMGA2 oncogene were associated with height. As far as associations, the research highlights that 'up to 90% of the variation in height among most human populations can be attributed to DNA' (Weedon et. al., p. 2: 2007). Our variable representing a child's own perception of her home environment may also serve as a control for a child's unobserved genetic expression as it is likely to be correlated with parents' response to child-specific characteristics.

important features of observed inputs. Probably most importantly, quality of parental investment is unobserved. As noted by Waldfogel, 'maternal sensitivity is the most important predictor of child social and emotional development' (Waldfogel, p. 62: 2006), which is unobserved.[4] To the extent that our observed measures for the latent variables capture such unobservables, the measurement model mitigates this problem.

Given that we only have data for two moments in time for each cohort, our analysis simplifies compared to that of Cunha and Heckman (2008). We are only able to estimate a single transition, for initially one-year- and eight-year-old children to age five and twelve respectively.[5] We therefore estimate a simple recursive system of equations:

$$\begin{pmatrix} \theta^C_{t-1} \\ \theta^N_{t-1} \end{pmatrix} = \Gamma_{t-1}\theta^I_{t-1} + \Upsilon_{t-1}X_{t-1} + \begin{pmatrix} \zeta^C_{t-1} \\ \zeta^N_{t-1} \end{pmatrix} \tag{3a}$$

$$\begin{pmatrix} \theta^C_{t} \\ \theta^N_{t} \end{pmatrix} = B_t \begin{pmatrix} \theta^C_{t-1} \\ \theta^N_{t-1} \end{pmatrix} + \Gamma_t\theta^I_t + \Upsilon_t X_t + \begin{pmatrix} \zeta^C_{t} \\ \zeta^N_{t} \end{pmatrix} \tag{3b}$$

We make the assumptions that $E(B) = 0$, $E(\Gamma) = 0$, $E(\Upsilon) = 0$, $Cov(B, \Gamma) = 0$, $Cov(B, \Upsilon) = 0$ and $Cov(\Upsilon, \Gamma) = 0$. While system (3) is relatively straightforward to estimate under these assumptions due to its recursive structure, the main problem arises from the fact that skills and parental investment are latent variables and have to be estimated given available indicators through confirmatory factor analysis. More specifically, given our objective to estimate (3), we are only 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

$$x^k_{i,t} = b^k_{io,t} + b^k_{i1,t}\theta^k_t + u^k_{i,t} \tag{4}$$

where x represents observed measures of the latent variable with $i = 1, ..., m^k_t$ denoting the different available indicators for the specific latent variable; θ is the factor for the latent variable k with $k \in \{C, N, I\}$ and u is an error term where θ and u are unobserved. b_{i1} represents factor loadings and $b^k_{io,t}$ 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,

[4] To some extent, as described in the data section later, our PRF variable captures such sensitivity as the indicators used include variables proxying parental care exerted on the child, such as breast-feeding or intensity of antenatal care. Unfortunately, we do not possess similar measures for five-, eight-, and twelve-year-old children.

[5] We also have to assume that initial conditions are the same across children of the older cohort as we do not have any information on their early childhood characteristics.

the errors are independent over time and across children. Thirdly, we assume that the relationship between the factor and the observed variables is linear.[6] Finally, the scale of the common factor is fixed by setting the first factor loading equal to one.[7] Equation (4) is estimated using Maximum Likelihood. The factor score is then predicted as the conditional mean of the latent variable given the observed variables.

The factor models for the latent skill variables are written in vector notation as follows

$$X_{i,t-1}^{C} = \Lambda_{oi,t-1} + \Lambda_{1i,t-1}^{C}\theta_{t-1}^{C} + \epsilon_{1i,t-1}^{C} \tag{5a}$$
$$X_{i,t-1}^{N} = \Lambda_{oi,t-1} + \Lambda_{2i,t-1}^{N}\theta_{t-1}^{N} + \epsilon_{2i,t-1}^{N} \tag{5b}$$
$$X_{i,t}^{C} = \Lambda_{oi,t} + \Lambda_{3i,t}^{C}\theta_{t}^{C} + \epsilon_{3i,t}^{C} \tag{5c}$$
$$X_{i,t}^{N} = \Lambda_{oi,t} + \Lambda_{4i,t}^{N}\theta_{t}^{N} + \epsilon_{4i,t}^{N} \tag{5d}$$

The factor models for the latent parental investment variables are

$$X_{i,t-1}^{I} = \Lambda_{oi,t-1} + \Lambda_{1i,t-1}^{I}\theta_{t-1}^{I} + \delta_{i,t-1}^{I} \tag{6a}$$
$$X_{i,t}^{I} = \Lambda_{oi,t} + \Lambda_{2i,t}^{I}\theta_{t}^{I} + \delta_{i,t}^{I} \tag{6b}$$

We allow latent cognitive and non-cognitive skill variables for the same age of a child to covary, and for parental investment indicators also across age. The covariance matrix for θ is therefore

$$\Phi = \begin{pmatrix} \phi_{11} & & & & & \\ \phi_{21} & \phi_{22} & & & & \\ 0 & 0 & \phi_{33} & & & \\ 0 & 0 & \phi_{43} & \phi_{44} & & \\ 0 & 0 & 0 & 0 & \phi_{55} & \\ 0 & 0 & 0 & 0 & \phi_{65} & \phi_{66} \end{pmatrix} \tag{7}$$

In contrast, we set all off-diagonal elements of the variance-covariance matrix for

[6]Cunha, Heckman, and Schennach (2006) analyse the more general case of non-linear systems.

[7]Alternatively, we could set the factor variance equal to one, which would result in an equivalent model as restricting the first factor loading to one. However, the latter restriction appears to be preferable from the point of view of factorial invariance (Rabe-Hesketh and Skrondal, 2004).

the errors of the factor models to zero, i.e., $Cov(\epsilon_t^q, \epsilon_t^p) = 0$, $Cov(\epsilon_t^q, \epsilon_{t-1}^p) = 0$ and $Cov(\epsilon_t^q, \epsilon_{t-1}^q) = 0$ for $q, p \in \{C, N, I\}$, which leaves us with two diagonal matrices.

$$\Xi_\epsilon = diag(\Xi_{t-1}^{\epsilon^C}, \Xi_{t-1}^{\epsilon^N}, \Xi_t^{\epsilon^C}, \Xi_t^{\epsilon^N}) \tag{8}$$

$$\Xi_\delta = diag(\Xi_{t-1}^{\delta}, \Xi_t^{\delta}) \tag{9}$$

While the assumption of diagonal error matrices for the factor models helps achieving identification of the model, it may appear restrictive - in particular with regard to parental investment. Finally, we allow for the possibility that the variance-covariance matrix Σ for the system (3) has off-diagonal elements not equal to zero for cognitive and non-cognitive skill levels at age eight (Equation 3a) and twelve (Equation 3b). By Choleski factorisation, we can estimate the following sample error covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_{11} & & & \\ \sigma_{21} & \sigma_{22} & & \\ 0 & 0 & \sigma_{33} & \\ 0 & 0 & \sigma_{43} & \sigma_{44} \end{pmatrix} \tag{10}$$

Before proceeding to the estimation of the model, we have to ensure that it is identified. As suggested by Jöreskog and Sörbom (1978), an estimation-based check for identification of the model is to look at the information matrix obtained from the maximum likelihood estimation. If the matrix is positive definite, Jöreskog and Sörbom suggest that the model is highly likely to be identified. It is preferable, however, to ensure identification before we proceed with the estimation step of the analysis. To prove identification, we have to identify the joint distributions of $\{\theta_{i,t-1}^C, \theta_{i,t-1}^N\}_{t-1}$, $\{\theta_{i,t}^C, \theta_{i,t}^N\}_t$, $\{\theta_{i,t-1}^I\}_{t-1}$, $\{\theta_{i,t}^I\}_t$, $\{\epsilon_{i,t-1}^k\}_{t-1}^i$, $\{\epsilon_{i,t}^k\}_t^i$, $\{\delta_{i,t-1}^I\}_{t-1}^i$, $\{\delta_{i,t}^I\}_t^i$ for $k, l \in \{C, N\}$ and the factor loadings $\{\lambda_{i,t-1}^q\}_{t-1}^i$ and $\{\lambda_{i,t}^q\}_t^i$ for $q \in \{C, N, I\}$. In order to achieve this, we have to identify the variances and covariances of the latent variables by showing that all latent variables can be solved for in terms of the variances and covariances of the observed indicators. Once we have demonstrated that the parameters of the measurement model are identified in this way, the parameters of the recursive model can be identified by solving them in terms of the covariances among the factors (Long, 1983).

We have assumed above that covariances for the latent variables are set to zero for all off-diagonal elements of Φ except for latent variables for the same age, and parental investment latent variables also across age. Together with the assumption that Ξ_ϵ and Ξ_δ are diagonal matrices, this implies that most covariances are set to zero. We only have to ensure identification of the non-zero off-diagonal elements. In order to see that they are identified, we first show that the factor loadings are identified by dividing (11b) by (11a) and (12b) by (12a) which express covariances of latent variables in terms of covariances of observed variables where $k, l \in \{C, N\}$. We assume without loss of generality two observed indicators $m_t^q = m_{t-1}^q = 2$ with $q \in \{C, N, I\}$ for each latent variable (subscripts for factor loadings omit m^q for ease of presentation).

$$Cov(X_{1,t-1}^k, X_{1,t-1}^l) = Cov(\theta_{t-1}^k, \theta_{t-1}^l) \tag{11a}$$

$$Cov(X_{1,t-1}^k, X_{2,t-1}^l) = \lambda_{t-1}^k Cov(\theta_{t-1}^k, \theta_{t-1}^l) \tag{11b}$$

$$\frac{Cov(X_{1,t-1}^k, X_{2,t-1}^l)}{Cov(X_{1,t-1}^k, X_{1,t-1}^l)} = \lambda_{t-1}^k \tag{11c}$$

and similarly

$$Cov(X_{1,t}^k, X_{1,t}^l) = Cov(\theta_t^k, \theta_t^l) \tag{12a}$$

$$Cov(X_{1,t}^k, X_{2,t}^l) = \lambda_t^k Cov(\theta_t^k, \theta_t^l) \tag{12b}$$

$$\frac{Cov(X_{1,t}^k, X_{2,t}^l)}{Cov(X_{1,t}^k, X_{1,t}^l)} = \lambda_t^k \tag{12c}$$

Given the factor loadings in (11c) and (12c), the covariances across latent skills for the same age are identified, as they can be written as the ratio of the observed covariance of observed indicators and the identified factor loading.

$$Cov(\theta_{t-1}^k, \theta_{t-1}^l) = \frac{Cov(X_{1,t-1}^k, X_{2,t-1}^l)}{\lambda_{t-1}^k} \tag{13}$$

and

$$Cov(\theta_t^k, \theta_t^l) = \frac{Cov(X_{1,t}^k, X_{2,t}^l)}{\lambda_t^k} \tag{14}$$

where $k \neq l$. In case $k = l$, we proceed in a slightly modified way. First we note that

$$Var(X_{1,t-1}^k, X_{1,t-1}^k) = Var(\theta_{t-1}^k) + \epsilon_{t-1}^k \tag{15a}$$

$$Cov(X_{1,t-1}^k, X_{2,t-1}^k) = \lambda_{t-1}^k Var(\theta_{t-1}^k) \tag{15b}$$

and

$$Var(X^k_{1,t}, X^k_{1,t}) = Var(\theta^k_t) + \epsilon^k_t \tag{16a}$$
$$Cov(X^k_{1,t}, X^k_{2,t}) = \lambda^k_t Var(\theta^k_t) \tag{16b}$$

$Var(\theta^k_t)$ and $Var(\theta^k_{t-1})$ are identified as they are given by

$$Var(\theta^k_{t-1}) = \frac{\lambda^k_{t-1} Var(\theta^k_{t-1})}{\lambda^k_{t-1}} \tag{17}$$

and

$$Var(\theta^k_t) = \frac{\lambda^k_t Var(\theta^k_t)}{\lambda^k_t} \tag{18}$$

which implies that also ϵ^k_{t-1} and ϵ^k_t are identified. Finally, for the case of parental investment, where we allow latent variables to be correlated across age, identification is shown by

$$Cov(X^I_{1,t-1}, X^I_{1,t}) = Cov(\theta^I_{t-1}, \theta^I_t) \tag{19a}$$
$$Cov(X^I_{2,t-1}, X^I_{1,t}) = \lambda^I_{t-1} Cov(\theta^I_{t-1}, \theta^I_t) \tag{19b}$$
$$\frac{Cov(X^I_{2,t-1}, X^I_{2,t})}{Cov(X^I_{1,t-1}, X^I_{1,t})} = \lambda^I_{t-1} \tag{19c}$$

and similarly

$$Cov(X^I_{1,t-1}, X^I_{1,t}) = Cov(\theta^I_{t-1}, \theta^I_t) \tag{20a}$$
$$Cov(X^I_{1,t-1}, X^I_{2,t}) = \lambda^I_t Cov(\theta^I_{t-1}, \theta^I_t) \tag{20b}$$
$$\frac{Cov(X^I_{1,t-1}, X^I_{2,t})}{Cov(X^I_{1,t-1}, X^I_{1,t})} = \lambda^I_t \tag{20c}$$

and following the same logic as above,

$$Cov(\theta^I_{t-1}, \theta^I_t) = \frac{Cov(X^I_{2,t-1}, X^I_{1,t})}{\lambda^I_{t-1}} \tag{21}$$

and

$$Cov(\theta^I_{t-1}, \theta^I_t) = \frac{Cov(X^I_{1,t-1}, X^I_{2,t})}{\lambda^I_t} \tag{22}$$

and then to identify the variances and error terms

$$Var(X^I_{1,t-1}, X^I_{1,t-1}) = Var(\theta^I_{t-1}) + \epsilon^I_{t-1} \tag{23a}$$

$$Cov(X^I_{1,t-1}, X^I_{2,t-1}) = \lambda^I_{t-1} Var(\theta^I_{t-1}) \tag{23b}$$

$$Var(\theta^I_{t-1}) = \frac{\lambda^I_{t-1} Var(\theta^I_{t-1})}{\lambda^I_{t-1}} \tag{23c}$$

and

$$Var(X^I_{1,t}, X^I_{1,t}) = Var(\theta^I_t) + \epsilon^I_t \tag{24a}$$

$$Cov(X^I_{1,t}, X^I_{2,t}) = \lambda^I_t Var(\theta^I_t) \tag{24b}$$

$$Var(\theta^I_t) = \frac{\lambda^I_t Var(\theta^I_t)}{\lambda^I_t} \tag{24c}$$

where again errors are identified given that we have identified variances and covariances in (23c) and (24c). Hence, given that we have shown that the model is identified, we proceed with estimation.

3 Data

We use data from the India part of the Young Lives (YL) project. YL is a long-term study of childhood poverty being carried out in Ethiopia, India (in the state of Andhra Pradesh), Peru and Vietnam. The survey consists of tracking two cohorts of children over a 15-year period. Currently we are able to use information from two rounds of data collection. In Round 1, 2,000 children aged around one (the 'younger' cohort) and 1,000 children aged around eight (the 'older' cohort) were surveyed in 2002. Following up, Round 2 involved tracking the same children and surveying them in 2006 at age five and twelve respectively.

The sample of children is representative of the three regions of Andhra Pradesh: Rayalseema, Coastal Andhra and Telangana. The sampling process was fourfold. 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 approximately four to five villages that comprised a sentinel site. Finally the questionnaires were administered to around 100 one-year-old and 50 eight-year-old children in these villages. Data was collected through household questionnaires, child questionnaires and a community questionnaire. Our estimation incorporates this survey design, wherein we use regions as our stratification variable and the sentinel sites as our clustering variable.

We use data obtained from both cohorts of children available in the YL survey. The two cohorts allow us to investigate two distinct periods of childhood. During the early childhood years, the transition between age one and five, a child still depends fully on her parents and family. The first few years of a child's life are decisive for the child's later physical and psychological well-being. The child learns during these years above all how to self-regulate, i.e., how to control her attention, emotions and behaviours. At the same time, the child acquires crucial cognitive skills, above all in terms of language acquisition. Therefore, the data on these early childhood years allow us to analyse factors influencing the foundations of skill formation, paying particular attention to a child's physical condition and her home environment. The data on the eight- to twelve-year-old cohort provides information on school-age children. Children at that age are concerned with the development of reflection - both on their own and others' thinking - and begin to think ahead in time and make plans for their future assuming responsibility for their actions. This goes along with increased social and emotional awareness. It is an age at which children begin to realise that they live in a society which sets challenges for them and they start figuring out how to find their own position within it. Apart from increasing cognitive skills, it is important for children to build confidence during this time, as it will be crucial later when they become more independent. An important change in a child's life during that period is a shift in importance of interaction with parents to peer interaction. Yet, while peer interaction becomes an essential part of a school-age child's existence, parents remain at the center of the child's life. Their role now shifts towards regulating a child's behaviour through monitoring and discipline (Waldfogel, 2006). In particular, the data on older children allows us to specifically analyse the dynamics of cognitive and non-cognitive skill formation and influences exerted by the child's immediate environment.

3.1 Latent variable indicators

Given that cognitive and non-cognitive skills and parental input are in principle unobserved, we have to treat them as latent variables for which we need to find observed indicators. Our data set provides us with multiple measures of child cognitive and non-cognitive skills as well as measures for parental investment, for both cohorts. We note here that the survey questionnaires differed slightly between Round 1 and Round 2. Although there is significant overlap between questions asked, we find that some indicators differ between the two rounds. However, as our interest lies in estimating latent variables, i.e., cognitive/non-cognitive achievement and parental investment, we are able to identify indicators for each of these in both rounds. As a result, these indicators can be different for the same cohort between the two rounds, but essentially lend to measurement in similar ways.

We rely on a range of observed indicators, for each cohort, to estimate our latent variables of interest: cognitive ability, non-cognitive ability, parental investment, child health, and PRF. Table (A) provides a comprehensive list and description of variables that we use as indicators in the measurement model to measure the above-mentioned items. As indicators of child health we use anthropometry z-scores, i.e., weight for age and height for age and whether the child has suffered from serious illness.[8] For children aged one, we do not observe any measures of cognitive and non-cognitive skills and therefore rely on child stuntedness as our outcome of interest.

In construction of aggregate scores for cognitive outcomes, we make use of scores calibrated using Item Response Theory (IRT) instead of raw aggregates of all sub-items. Aggregation does not account for differences between the individual questions on the test: the probability of giving a correct response to some questions depends on a child's ability, the level of difficulty of that particular item and its discriminating power between high and low ability individuals. Using IRT, we specify a three parameter logistic model for the probability of a correct response by an individual to the different test questions. The three parameters are: item discrimination, item difficulty and the lower asymptote for the probability function denoting random guessing on an item. Using maximum likelihood we obtain the expected score on each section of the test as our outcome variable. We significantly reduce measurement error in estimating latent cognitive ability by this procedure.

3.2 Home environment and other input measures

Further, we have information on a range of home inputs, child characteristics and caregiver characteristics for both cohorts for both survey rounds. We use, as measures for home environment, various household level attributes: household size, location (whether urban or rural), primary occupation (whether non-agricultural/salaried or

[8] For the older cohort we substitute child weight and height z-scores with body mass index z-scores.

other), social networks (given by indicators such as number of groups a household is affiliated to, community based participation of household, kinship ties within the community of household), household mean educational attainment, caste, and asset ownership.[9] We also classify households as poor/non-poor using monthly per-capita expenditure information based on a poverty line of Rs. 292.95 and Rs 542.89 for rural and urban areas respectively in the state of Andhra Pradesh (GOI, 2007).[10] For child characteristics, we use information on long-term health of a child (child weight and height for the 'younger' cohort), number of siblings, gender, and number of years of schooling attained. We include also child's own perception of well-being at home. This variable factors many questions pertaining to child overall happiness and satisfaction with family and household. Additionally, we account for the type of school (public or private) that the child is enrolled in. Finally, the YL questionnaires explicitly identify and administer questionnaires to the caregiver of the child, which allows us to include education of the caregiver as a variable of interest. In addition, we construct from the information in the caregiver questionnaire a variable reflecting the degree of parental altruism towards the child. A common assumption in the existing literature is that the only motivation for parents to raise a child is altruism (Cunha and Heckman, 2007). Yet, it is a well-known fact that this assumption is unlikely to hold in an economically deprived environment such as Andhra Pradesh where children serve both as a source of current income and means of providing income when parents are too old to work. Nevertheless, we believe that also parents in deprived environments ideally want to see their children grow to become healthy and happy individuals. In order to reflect this tension between parental interests and control for effects from varying degrees of altruism across the sample, we constructed a variable reflecting parents' motivation to have the YL child. This information is only available for Round 2 data.

4 Estimation

Beginning with the younger cohort, we note that, better child health or lack of child stuntedness is known to be associated with later period skill accumulation (Grantham-McGregor et al., 2007; Walker et al., 2007). Here, our objective is to establish links between childhood poverty, PRF and early child health, where PRF is treated as a latent variable which is predicted by the indicators described in Table (A). Subsequently we

[9]The household questionnaire, recorded the ownership of each asset (about 20 different types) present in the household of the child. We were able to combine this information into an asset index by the method of principal components.

[10]Household consumption expenditure comprises the expenditure on food, non-food items and consumer durables. Expenditure per capita calculations take into account the age/sex specific equivalence factors.

test whether early childhood health impacts cognitive/non-cognitive outcomes at age five. Hence, for the 'younger' cohort, our structural model (3) simplifies as (3a) contains only observed child health as the dependent variable, which is the only period $t-1$ variable feeding into (3b). Instead of parental investment for one-year-old children in (3a), latent PRF is used to reflect parental input. For the 'older' cohort, Table (A) shows that we possess sufficient indicators to estimate the full measurement model consisting of (5) and (6). The estimated factors are used in the estimation of the complete structural model (3). It is therefore the 'older' cohort that allows us to investigate the presence and importance of self- and cross-productivity of skill levels.

As discussed in the previous section, in estimating the factor models, we set $\lambda^k_{o,t} = 1$ with $k \in \{C, N\}$ and $\lambda^I_{o,t} = 1$ to fix the scales of the measurement in the latent variables. When including the exogenous covariates, there is one disadvantage in using LISREL, which we employ for our estimation. The explanatory variables have to be introduced into the model as artificial latent variables by setting their factor loadings equal to one and zero for all other observed covariates. Moreover, the unique factor variance is set to zero. This implies that the explanatory variables are treated as response variables for which multivariate normality is assumed. We estimate (5) and (6) without intercepts as we use mean centered observed indicators $X^k_{i,t}$ with $k \in \{C, N\}$ and $X^I_{i,t}$.[11] In principle, the recursive model could be estimated using OLS if we assume that the latent variables are independent of the error term ζ^q_t for $q \in \{C, N, I\}$. More specifically, (3a) could be consistently estimated applying OLS under the assumption that $Cov(\theta^I_{t-1}, \zeta_{t-1}) = 0$. Then also (3b) could be consistently estimated using OLS under the additional assumption that not only $Cov(\theta^I_t, \zeta_t) = 0$ but also $Cov(\theta^k_{t-1}, \zeta_t) = 0$ for $k \in \{C, N\}$. This is the case since we have assumed in (10) that $Cov(\zeta_t, \zeta_{t-1}) = 0$. This means that we can regard θ^k_{t-1} as statistically predetermined with regard to θ^k_t. Obviously, this requires us to assume that errors are serially uncorrelated, i.e., $Cov(\theta^k_{t-1}, \zeta_t) \neq 0$.[12] Instead of using OLS, we estimate the system (3) using Full Information Maximum Likelihood (FIML). FIML not only provides us with the most efficient estimators of system (3), it also tackles the problem of missing data, a common occurrence with perception based response questions. Using FIML, enabling listwise deletion, we are able to integrate out the missing data from the sample likelihood.

Finally, an important contribution of the analysis of Cunha and Heckman (2008) was to anchor scores of skills in a child's adult earnings. We lack such information on a child's eventual success in life which impedes anchoring of cognitive and non-cognitive

[11] The results of estimating the factor models are reported in the Appendix but are not discussed in Section 5.

[12] This assumption is also made by Cunha and Heckman (2008).

scores in order to generate a cardinal measure of skills. Instead, we have to rely on the ordinal measures our data generates. This implies that we can compare a child's performance only relative to the other children in the sample. In a cross-section, this is unproblematic; however, ideally we would like to be able to verify whether a child has advanced in terms of her skill levels over time. This is not possible in absolute terms; we can only measure whether a child has moved up in the relative ranking vis-a-vis the other children in the sample.

5 Results

Because we are mainly interested in investigating the presence of self-productivity and cross-productivity, we focus in our graphical analysis on the 'older' cohort. Box-plots 1-4 show the dynamic relationships for the 'older' cohort between a child's cognitive and non-cognitive skills. More specifically, the plots for the 'older' cohort draw the unconditional relationship between the distribution of children's skill levels - cognitive or non-cognitive - at age eight *by decile* against the distribution of skill scores at age twelve. Using box-plots by decile is robust to children that have made extreme jumps in terms of their skill development, either positively of negatively, and therefore reveals the general pattern of skill formation during the periods of transition from age one to five and eight to twelve.

Examining these plots for the 'older' cohort allows us to look for preliminary evidence of self- and cross-productivity. Box-plot 1 demonstrates that children with a lower score for cognitive skills at age eight also tend to have lower scores for cognitive skills at age twelve. This represents pronounced evidence in favour of the presence of self-productivity for the formation of cognitive skills. For non-cognitive skills, the unconditional relationship between the non-cognitive skill level for eight-year-old and twelve-year-old children also seems to be characterised by positive correlation, but self-productivity appears to be less pronounced than for cognitive skills. Yet, this also points to the presence of self-productivity for non-cognitive skills. Analysing the relation between the non-cognitive skill level at age eight and the cognitive skill level at age twelve in box-plot 3 also shows clear evidence of the presence of cross-productivity. Similarly, children with lower cognitive skill levels at age eight tend to have lower non-cognitive skill levels at age twelve. Hence, the graphical analysis points to the presence of self-productivity and cross-productivity for both cognitive and non-cognitive skills.

The box-plots only plot the unconditional relation between skill levels. Figures 5-8 in contrast, plot the conditional probability density of the skill level for a twelve-year-old child on her skill level at age eight $\hat{f}(\theta_t^k|\theta_{t-1}^k)$ with $k \in \{C, N\}$. The density is estimated nonparametrically using the simple '$1.06\hat{\sigma}n^{-1/5}$' rule-of-thumb bandwidth based on a Gaussian Kernel with $\hat{\sigma}$ being the sample standard deviation and n the number of observations (Li and Racine, 2007).[13] Figure 5 confirms the impression obtained from the corresponding box-plot 1. With the probability density mass concentrated along the diagonal, there is strong evidence that a child has a higher cognitive skill level at age twelve conditional on her having achieved a high cognitive skill level at age eight. Moreover, the relationship appears to be linear, providing support to our linear specification of the system in (3). In contrast, the conditional probability density plot 6 for non-cognitive skills is less clear-cut. The probability density mass appears to be larger for lower values of non-cognitive skills for children aged eight as well as twelve. This suggests that a child has lower non-cognitive skills at age twelve conditional on having had a lower non-cognitive achievement score at age eight. Yet, this kind of relationship is less evident for children with higher non-cognitive achievement. The density plots examining cross-productivity 7 and 8 have a similar shape as the non-cognitive skill plot 6. Plotting the density of cognitive skill levels at age twelve conditional on non-cognitive skill levels at age eight 7, shows that lower non-cognitive achievement at age eight is associated with a relatively large range of cognitive skill levels at age twelve. This points to a possibly weak relationship between non-cognitive skill levels at age eight and cognitive achievement at age twelve, a finding which we will confirm below. Moreover, the plot suggests a possible non-linear relation between non-cognitive skills at age eight and cognitive skills at age twelve. Plot 8 of non-cognitive skills at age twelve conditional on cognitive skills at age eight points more clearly to the presence of cross-productivity with a probability mass more closely aligned along the diagonal. In brief, the conditional probability density plots confirm the overall associations drawn by the unconditional distribution box-plots.

Table 5 shows the pairwise Spearman rank correlation matrix of the predicted latent variables.[14] For the 'younger' cohort, the correlation matrix contains correlations for cognitive, non-cognitive skills, parental investment at age five, a child's health condition and a measure of psycho-social risk factors. This measure of psycho-social risk factors is negatively correlated with a child's health condition. It is also negatively correlated with cognitive and non-cognitive skill levels as well as parental investment at age five. Child health at age one itself is strongly positively correlated with cognitive and non-cognitive abilities, suggesting early-age health has important implications for later skill formation. Finally, the table also shows that higher levels of parental investment are associated with higher levels of both cognitive and non-cognitive skills. For the 'older'

[13] We use the *np* package in R to estimate the conditional density (Hayfield and Racine, 2008).

[14] Rank correlation coefficients are more robust to the presence of outliers and therefore appear in our setting preferable over Pearson's product-moment correlation coefficient.

cohort, the matrix shows correlations of the skill variables and parental investment for both age eight and twelve. The matrix confirms our graphical findings that there exists a positive association between a child's skill level at age eight and her eventual skill level at age twelve, both within the same skill and across skills. The extremely high correlation coefficient for cognitive skills at age eight and twelve is particularly interesting as it suggests a very close relation between a child's cognitive ability level at age eight and age twelve. Similar to the younger cohort, we find the parental investment variable for eight- as well as twelve-year-old children to be positively correlation with skill levels contemporaneously and over time. Therefore, the correlation matrix provides further evidence in favour of the presence of both self- and cross-productivity.

Before we proceed with a discussion of the results of estimating the structural model (3), Tables 1 and 2 contain summary statistics for all the variables used in the structural model. While many variables, such as gender or the number of siblings, have a straightforward and intuitive meaning, interpreting variables that we have constructed is much less straightforward. For example our variable proxying social networks, HH Social Connectedness, has no direct intuitive interpretation, apart from a purely directional interpretation that 'more is better'. The same applies for the wealth index that we have created to capture the entirety of a household's wealth. Tables 3 and 4 provide summary statistics of the indicator variables used in the measurement models (5) and (6).

We first report results for the younger cohort. From Table 6 we see that, both household poverty and psycho-social risk factors exert a statistically significant influence on child health at age one. Both household poverty and high psycho-social risk factors adversely affect child health, causing child stuntedness. Given this, we go on to establish the effect of child health on learning outcomes for the child aged five. We find that child health positively affects child cognitive skills. Its effect on child non-cognitive skills is not significant. These results are consistent with the literature discussed in the Introduction that link early child stuntedness with later period skill development and attribute the lack of development in the child's early stages to certain home environment factors, i.e., PRF and poverty. We also find a positive effect of household education on child learning. Well-educated parents produce well-educated offspring, a fact well established in many previous studies: 'one of the important roles that parents play in their child's development has to do with the stimulation of cognitive and language growth in the first few years of life' (Waldfogel, p. 49: 2006). Also, our measure of a household's social network shows a positive statistically significant effect on cognitive skills levels, while the effect is negative for non-cognitive skills.[15]

Next, we discuss the results for the 'older' cohort. For children of age eight, we see from Table 7, that parental investment has a significant positive effect on cognitive skill accumulation but an insignificant effect on non-cognitive skill accumulation, while the sign is still positive. Apart from this, we find additionally that child health and the number of siblings affect child cognitive outcomes positively. Social networks exert a positive effect on cognitive skills, similar to our findings for five-year-old children. This is consistent with other findings that social capital both within the family, and even more so within the community, has a positive effect on children's educational achievements (Croll, 2004). We find that children attending private *elementary* schools perform much better on achievement tests compared to their public school counterparts, a result well-established for India in previous research studies (Kingdon, 1996; Muralidharan and Kremer, 2006). For non-cognitive skills, we find significant positive effects from households asset ownership and belonging to a household located in an urban area and negative effects from larger households, possibly reflecting a lack of parental attention to the child.

Finally, we examine self- and cross-productivity effects. Table 8 reports results of the 'older' cohort at age twelve. The results show that cognitive skills acquired at age eight, affect positively both cognitive and non-cognitive skills at age twelve. Thus we find evidence for the dynamic persistence of cognitive skills as also of its contribution towards developing non-cognitive skills at a later stage (as cross-productivity effects). In contrast, we find no such evidence for non-cognitive skills acquired at age eight. Its effect on both cognitive and non-cognitive skills at age twelve is insignificant. We find however substantial evidence in favour of parental investment, which exerts a strong significant and positive effect on cognitive and non-cognitive skill acquisition at this age. This effect is much larger in magnitude for cognitive achievement as compared to non-cognitive achievement. Interestingly, at this age we find that children attending a *public secondary* school have higher cognitive and non-cognitive achievement scores. This is in sharp contrast to the results for eight-year-old children, who perform better when attending a *private primary* school. One reason, perhaps, is that the growth and spread of private schools is confined to primary schooling, its presence remaining remarkably scarce within the secondary schooling sector. Moreover government subsidies

[15]The latter effect is quite counterintuitive. This is perhaps because, we use as one of ours indicator for non-cognitive skills, how well the child fares at pre-school (if we believe that children usually learn important social skills at pre-school rather than academic lessons). A child who does not enroll in pre-school, misses out on social interaction amongst peers and thus is given a value 0. As a result, it may be that parents who have a strong kinship network, choose to not send their children to preschool and instead rely on their social capital networks to raise their child.

are abundant in better quality public secondary schools (Kingdon, 1996). Much of the research on private schooling in India and other developing countries has focused on elementary schools. Further research into the nature and effects of private schooling for secondary schools is required to understand its implications. Next, we find that children belonging to wealthier households and households engaging in non-agricultural occupations show higher cognitive and non-cognitive skills, but belonging to a household engaged in a non-wage/agricultutral occupation lowers cognitive and non-cognitive achievements. Social connectedness is statistically significantly positive for cognitive skills, inline with our findings for children aged five and eight. Of special interest is that we find child self-reported well-being to contribute to better non-cognitive outcomes, reflecting the importance of a 'happy home' for successful non-cognitive development.

One of the contributions of Cunha and Heckman (2007) is to carve out sensitive and critical periods for skill formation during childhood. In our case, we have data only for a single transition for each cohort, from age one to five, and from age eight to twelve, which does not allow us to compare transitions over time for a single cohort. This is a disadvantage with regard to determining sensitive periods for skill formation. The concept of critical periods, however, does not seem to be uncontroversial, if we understand critical in the sense that a child must be exposed to a certain experience during a specific period of early childhood in order to fully acquire a certain skill. Waldvogel (2006) argues that this is generally not the case for skill formation in children. Only for very few competencies relating to some aspects of language acquisition require specific input during a well-determined period of early childhood. Hence, we regard our inability to determine critical periods as less important.

6 Insights

Recent research in economics has demonstrated the importance of early childhood for the development of the crucial skill set necessary for socio-economic success in later life. This research has shown that the acquisition of cognitive and non-cognitive skills begins from the very nascent stages of childhood and that these skills interact throughout a child's development process. These insights motivate the investigation of the determinants of these skills beyond formal schooling focusing in particular on self- and cross-productivity effects. So far, this research has been almost exclusively focused on children in industrialised countries.

Extending this research to the developing country context promises important insights with regard to finding answers to the question as to what constitutes an 'enabling' environment for a child's successful development. To this purpose, there are valuable lessons to be learnt from firstly exploring and ascertaining the link between cognitive and non-cognitive achievement and analysing this linkage in the context of a child's (immediate) environment in a developing country. Yet, conducting this kind of research in a developing country context is in many ways challenging as children face a drastically different environment to their peers in industrialised countries. Nevertheless, building on the recent contributions of Cunha and Heckman (2007, 2008) and exploiting a novel rich data set covering different periods of childhood, we estimate a structural model to investigate skill formation of children in Andhra Pradesh. Most importantly, for our 'older' cohort, we find evidence for the presence of self-productivity for cognitive skills and cross-productivity of cognitive on non-cognitive skills during the transition from eight to twelve years. We also find statistically significant evidence for parental investment, as measured by our latent variable, to contemporaneously exert an economically powerful positive influence on skill formation for five-, eight- and twelve-year-old children. Our results also point to a large number of other important determinants of skill formation including child, caregiver and household characteristics and notably also school type. The data available for our 'younger' cohort allows us to investigate the importance of earliest parental care and child health as well as the importance of child health achieved at age one for skill levels at age five. We find that so called psycho-social risk factors have a statistically significant effect on child health at age one. These factors are reflected in parental care during pregnancy as well as during the first few months of a child's life. Considering that we also find child health as measured for children at age one to represent a statistically significant determinant for a child's cognitive ability at age five, this provides powerful evidence to shift attention to providing parents with support from the earliest days of pregnancy onward.

References

[1] Andrabi, Tahir, Jishnu Das, Asim Khwaja and Tristan Zajonc (2008) *Do Value-Added Estimates Add Value? Accounting for Learning Dynamics*, BREAD Working Paper 170, Cambridge, MA: Harvard University

[2] Becker, G. and T. Nigel (1986) 'Human Capital and the Rise and Fall of Families', *Journal of Labor Economics* 4: 1-39

[3] Blau, D. and Janet Currie (2006) 'Preschool Daycare, and Afterschool Care: Who's Minding the Kids?' in E. Hanushek and F. Welch (eds.) *Handbook of the Economics of Education*, Amsterdam: North Holland

[4] Browning, Martin and J. Carro (2006) *Heterogeneity and Microeconometrics Modelling*, CAM Working Papers 2006-03, Copenhagen: University of Copenhagen, Department of Economics, Centre for Applied Microeconometrics

[5] Carneiro, P., Flavio Cunha and James Heckman (2003) 'Interpreting the Evidence of Family Influence on Child Development', paper presented at conference on The Economics of Early Child Development: Lessons for Economic Policy, Federal Reserve Bank, University of Minnesota, USA, 17 Oct 2003

[6] Carneiro, P. and James Heckman (2003) 'Human Capital Policy', in James Heckman, A. Krueger and B. Friedman (eds.) *Inequality in America: What Role for Human Capital Policies?*, Cambridge, MA: MIT Press

[7] Croll, P. (2004) 'Families, Social Capital and Educational Outcomes', *British Journal of Educational Studies*, 52.4: 390-416

[8] Cunha, Flavio and James Heckman (2006) 'Investing in our Young People', mimeo, University of Chicago, Department of Economics

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

[10] Cunha, Flavio and James Heckman (forthcoming) 'Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation', *Journal of Human Resources*

[11] Cunha, Flavio, James Heckman, L. Lochner and D. Masterov (2006) 'Interpreting the Evidence on Life Cycle Skill Formation', in Eric Hanushek and F. Welch (eds.) *Handbook of the Economics of Education*, Amsterdam: North Holland

[12] Cunha, Flavio, James Heckman and S.M. Schennach (2006) 'Estimating the Technology of Cognitive and Noncognitive Skill Formation', mimeo, Univerity of Chicago, Department of Economics

[13] Currie, Janet and D. Thomas (1995) 'Does Head Start Make a Difference?', *American Economic Review* 85.3: 341-364

[14] De Silva, M., S. Huttly, T. Harpham and M. Kenward (2007) 'Social Capital and Mental Health: A Comparative Analysis of Four Low Income Countries', *Social Science & Medicine* 64: 5-20

[15] Fafchamps, M. and F. Gubert (2006) 'The Formation of Risk Sharing Networks', *Journal of Development Economics* 83.2: 326-50

[16] Galab, S., H. Moestue, P. Anthony, A. McCoy, C. Ravi and P.P. Reddy (2006) *Child Learning in Andhra Pradesh: The Interplay Between School and Home*, Young Lives Working Paper, Oxford: Young Lives

[17] Government of India (GOI) (2007) 'Poverty Estimates For 2004-05', New Delhi: Planning Commission Report

[18] Grantham-McGregor, S., Y. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp and The International Child Development Steering Group (2007) 'Developmental Potential in the First 5 Years for Children in Developing Countries', *The Lancet* 369: 60-70

[19] Hayfield, T. and J. Racine (2008) 'Nonparametric Econometrics: The np Package', *Journal of Statistical Software* 27.5: 1-32

[20] Heckman, James (1995) 'Lessons from The Bell Curve', *Journal of Political Economy* 103.5: 1091-1120

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

[22] Heckman, J, J. Stixrud, and Urzua Sergio (2006) 'The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior', *Journal of Labor Economics* 24.3: 411-83

[23] Hunter, P. (2008) 'What Genes Remember', *Prospect* 146

[24] Jöreskog, K. and D. Sörbom (1981) 'LISREL VI Analysis of Linear Structural Relations by Maximum Likelihood and Least Squares Methods', Chicago: National Educational Resources

[25] Kingdon, Geeta (1996) 'Private Schooling in India: Size, Nature and Equity Effects', *Economic and Political Weekly* 31.51: 3306-14

[26] Li, Q. and J. Racine (2007) *Nonparametric Econometrics*, Princeton, NJ: Princeton University Press

[27] Long, Scott (1983) *Covariance Structure Models*, Sage University Paper No. 07-034, London: Sage Publications

[28] Moestue, H., S. Huttly, L. Sarella and S. Galab (2006) 'The Bigger the Better - Mothers' Social Networks and Child Nutrition in Andhra Pradesh', *Public Health Nutrition* 10.11: 1274-82

[29] Muralidharan, K. and M. Kremer (forthcoming) 'Public and Private Schools in Rural India', School Choice International, Conference Volume

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

[31] Scarr, Sandra (1992) 'Developmental Theories for the 1990s: Development and Individual Differences' *Child Development* 63.1: 1-19

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

[33] Skrondal, A. and S. Rabe-Hesketh (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*, FL: Chapman & Hall/CRC

[34] Teasdale, T. and K. Owen (1983) 'Heredity and Familial Environment in Intelligence and Educational Level - A Sibling Study', *Nature* 309: 620-22

[35] Thompson, R. and C. Nelson (2001) 'Developmental Science and the Media: Early Brain Development', *American Psychology* 56: 5-15

[36] Turkheimer, E., A. Haley, M. Waldron, B. D'Onofrio and I. Gottesman (2003) 'Socioeconomic Status Modifies Heritability of IQ in Young Children', *Psychological Science* 14.6: 623-28

[37] Todd, P.E. and K.I. Wolpin (2005) 'On the Specification and Estimation of the Production Function for Cognitive Achievement', *Economic Journal* 113.485: 3-33

[38] Todd, P.E. and K.I. Wolpin (2006) 'The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps', mimeo, University of Pennsylvania

[39] Waldfogel, Jane (2006) *What Children Need*, Boston: Harvard University Press

[40] Walker, S.P., T.D. Wachs, J. Gardner Meeks, B. Lozoff, G.A. Wasserman, E. Pollitt and J.A. Carter (2007) 'Child Development: Risk Factors for Adverse Outcomes in Developing Countries', *The Lancet* 369.9556: 145-57

[41] Weedon, M.N., G. Lettre, R.M. Freathy, C.M. Lindgren, B.F. Voight, J.R. Perry, K.S. Elliott, R. Hackett, C. Guiducci and B. Shields (2007) 'A Common Variant of Hmga2 is Associated with Adult and Childhood Height in the General Population', *Nature Genetics* 39.10: 1245-50

[42] Wilson, R.A. (1983) 'The Louisville Twin Study: Development Synchronies in Behavior', *Child Development* 54: 198-216

Table A : Measurement Model Indicators

	Cognitive Ability	Non-cognitive Ability	Parental Investment
<u>Age 8:</u> Older Cohort, Round 1	• Scores from the Raven's Progressive Matrices Aptitude Test • Reading Level • Writing Level	Child Mental Ability Indicators from the Strengths and Difficulties Questionnaire (SDQ) devised by Dr. Robert Goodman • Emotional Conduct • Hyperactivity • Pro-Social Behaviour • Conduct Problems	• Do parents spend money on child's education? • Does child work at home or outside? • How many years back was child made to start formal schooling? • How often does child see father (daily, weekly, once a month, once a year)
<u>Age 12:</u> Older Cohort, Round 2	• IRT scores from the Peabody Picture Vocabulary Test • Numeracy: child can perform simple multiplication • Writing Level	Child Personality Measures: indicated from questions rated on the Likehart Scale by child • Friendliness • Pride • Determination • Social Trust • Group Membership	• Proportion of total household expenditure on following received by child: o Education o Health o Clothing • How often does child see father (daily, weekly, once a month, once a year)
<u>Age 1:</u> Younger Cohort, Round 1			Psychosocial Risk Factors: • Caregiver education • No doctor present at birth? • Months left without breastfeeding • Level of antenatal care • Unwanted pregnancy? • Caregiver depression
<u>Age 5:</u> Younger Cohort, Round 2	• IRT scores from the Peabody Picture Vocabulary Test • Scores from the Cognitive Development Assessment - Quantitative Test	• Level of fluency and communication in native language • Performance in pre-school (interactive and social nature) • Does child travel to school with friends, parents or alone?	• Proportion of total household expenditure on following received by child: o Education o Health o Clothing • How often does child see father (daily, weekly, once a month, once a year)

Figure 1: Cognitive (Age 8) vs Cognitive (Age 12) Skill Levels

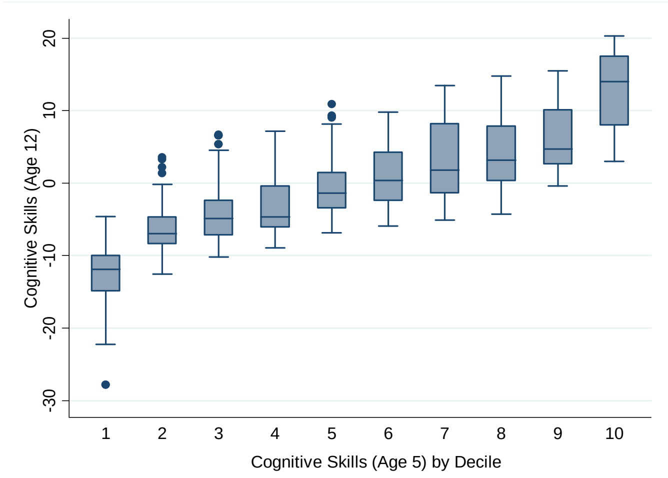

Figure 2: Non-cognitive (Age 8) vs Non-cognitive (Age 12) Skill Levels

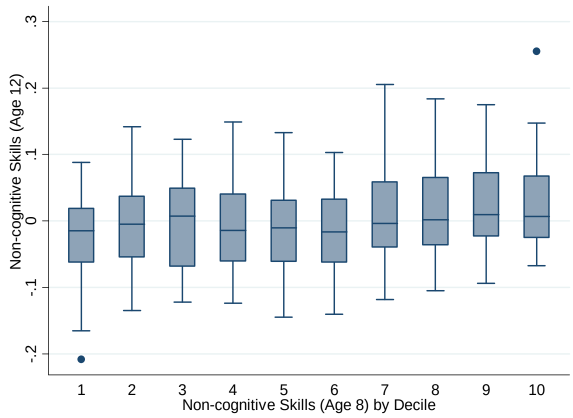

Figure 3: Non-cognitive (Age 8) vs Cognitive (Age 12) Skill Levels

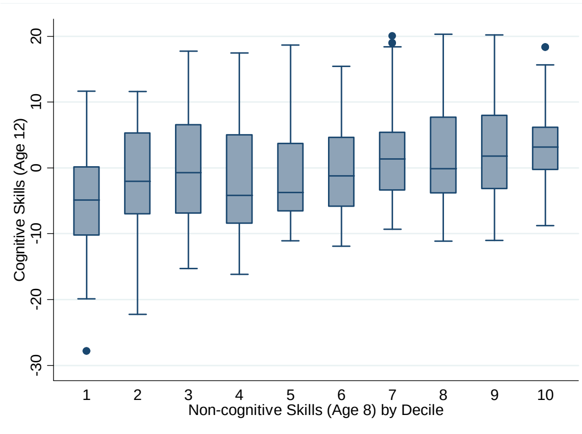

Figure 4: Cognitive (Age 8) vs Non-cognitive (Age 12) Skill Levels

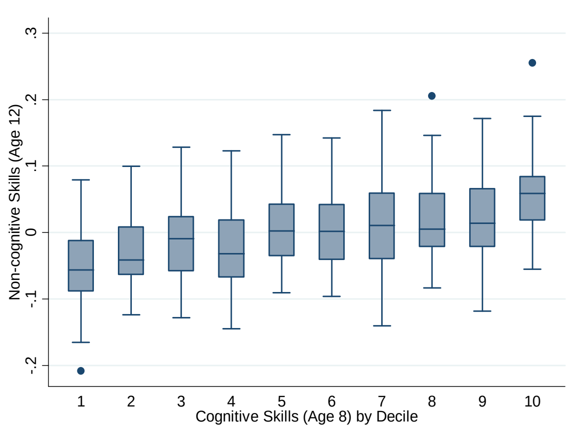

Figure 5: Cognitive (Age 8) vs Cognitive (Age 12): Conditional Probability Density Function

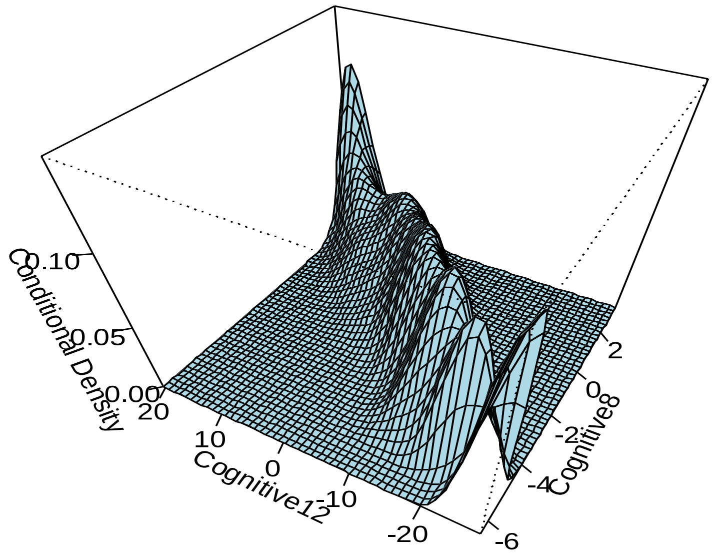

Figure 6: Non-cognitive (Age 8) vs Non-cognitive (Age 12) Skill Levels: Conditional Probability Density Function

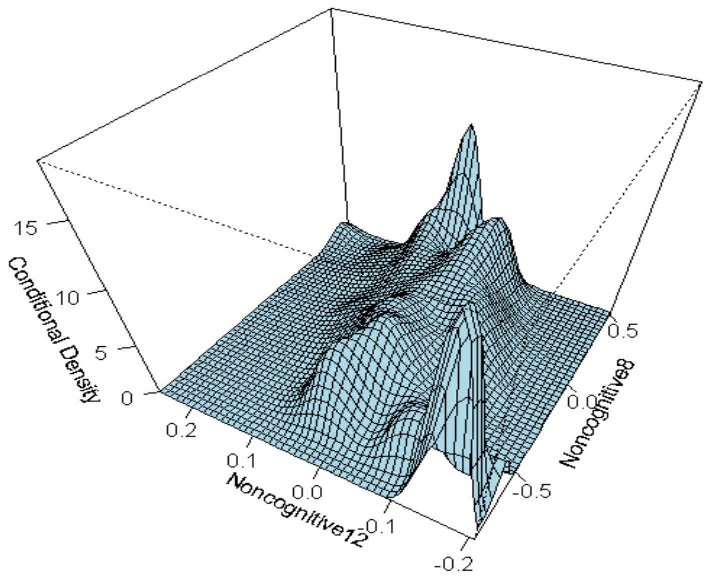

Figure 7: Non-cognitive (Age 8) vs Cognitive (Age 12) Skill Levels: Conditional Probability Density Function

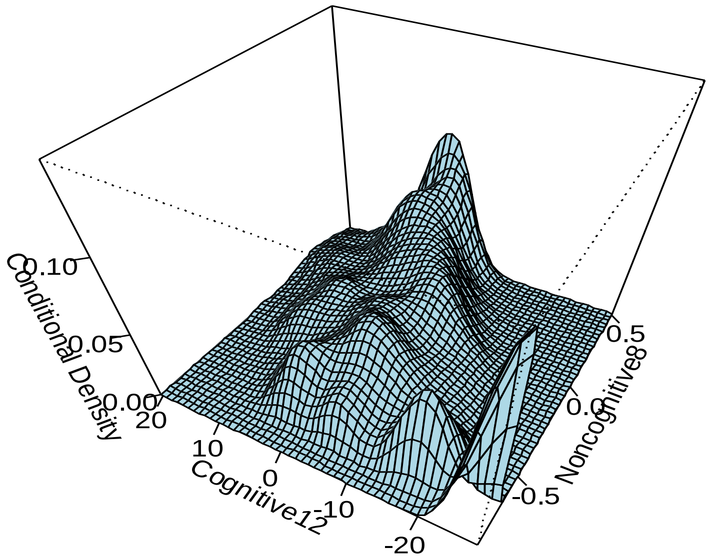

Figure 8: Cognitive (Age 8) vs Non-cognitive (Age 12) Skill Levels: Conditional Probability Density Function

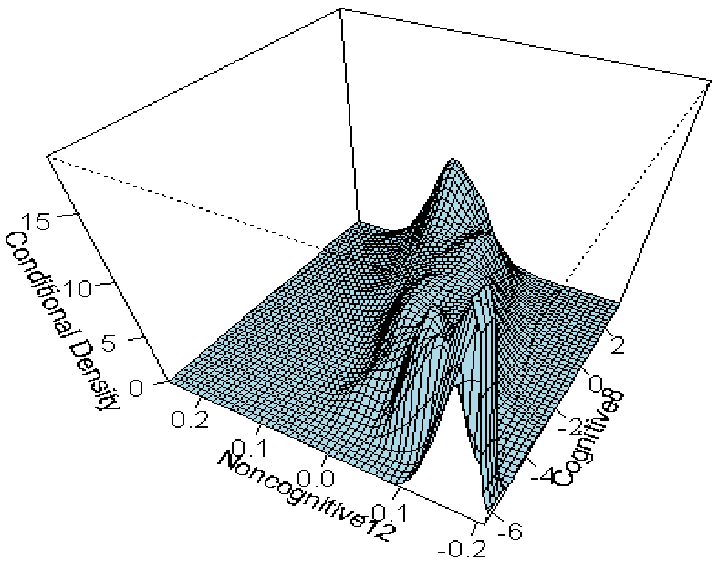

Table 1: Summary Statistics: Age 1 and 5

Table A: Age 1

Variable	No. Obs.	Median	Mean	St. Dev.
PRF[2]	1950	0.0164	0.245	3.04
Child Health[2]	1950	0.037	1.174	0.554
Poor	1914	0.692	0.698	0.092
CH Male	1950	0	0.463	0.499
CH Caste[3]	1950	0	0 .205	0.404
CH Birth Order	1936	1	1.638	0.956
HH Size	1950	5	5.425	2.369
HH Urban	1950	0	0.244	0.429
HH Non Ag. Occupation	1950	0	0.615	0.905
HH Mean Education	1950	8	7.018	5.004
HH Social Connectedness[4]	1950	-0.208	0.006	0.482

Table B: Age 5

Variable	No. Obs.	Median	Mean	St. Dev.
Cognitive Skills[2]	1950	-0.758	0.489	22.45
Non-cognitive Skills[2]	1950	-0.38	-0.018	1.29
Parental Investment[2]	1950	-0.009	-0.007	1.06
Parent Altruism[4]	1950	0	0.463	0.499
CH Siblings	1950	1	1.443	1.006
CG Education	1950	0	3.643	4.498
HH Size	1950	5	5.516	2.225
HH Social Connectedness[5]	1950	-.058	.0003	.654
HH Assets	1950	.361	.388	.211

Notes:
1. CH: Child; CG: Caregiver; HH: Household.
2. Estimated latent variables.
3. Indicator of whether child belongs to a non-SC/ST and non-backward Castee community.
4. Normalised on a scale of 0 to 1; is a combination of the responses to the questions on why parents feel it is important to have children.
5. An index (based on factor analysis) of the household's responses to questions of its kinship support base in community, how active it is (membership in societies etc.), level of trust within community.

Table 2: Summary Statistics: Age 8 and 12

Table A: Age 8

Variable	Obs.	Median	Mean	Std. Dev.
Cognitive Skills[2]	994	0.085	0.042	1.408
Non-cognitive Skills[2]	994	0.055	0.004	0.196
Parental Investment[2]	994	0.007	-0.000	0.066
Child Health[2]	994	0.010	0.002	0.223
CH Well-being[4]	994	0.666	0.697	0.227
CH Siblings	994	2	1.813	1.184
CH Male	994	1	0.509	0.500
CH Caste[5]	994	0	0.216	0.412
Public School	994	1	0.767	0.423
CG Education	994	4.605	4.628	0.037
HH Size	994	5	5.548	2.039
HH Social Connectedness[6]	994	-0.033	0.0004	0.488
HH Assets	994	0.309	0.340	0.212
HH Urban	994	0	0.242	0.429
HH Wage	994	0	0.606	0.816
HH Mean Education	994	6	6.662	4.246

Table B: Age 12

Variable	Obs.	Median	Mean	Std. Dev.
Cognitive Skills[2]	994	-0.633	0.117	8.03
Non-cognitive Skills[2]	994	-0.003	-0.000	0.064
Parental Investment[2]	994	-0.006	0.001	0.06
Child Health[2]	994	0.045	0.007	0.331
Parent Altruism[3]	994	0.701	0.702	0.092
CH Well-being[4]	994	0.791	0.783	0.125
CCH Height	985	141.2	140.845	11.61
CH Weight	985	30.75	32.33	10.65
CH Caste[5]	993	0	0.206	0.405
CH No. of Years of Schooling	980	6	5.601	1.263
CH School Starting Age	987	5	5.043	0.711
Public School	994	1	0.628	0.484
CG Educational Attainment	993	0	2.709	4.033
HH Size	994	5	5.197	1.832
HH Social Connectedness[6]	992	-0.024	0.002	0.669
HH Assets	994	0.378	0.395	0.214
HH Urban	994	0	0.251	0.434
HH Mean Education	994	9	9.958	5.707
HH Wage Recipient	994	0	0.569	0.809

Notes:
1. CH: Child; CG: Caregiver; HH: Household.
2. Estimated latent variables.
3. Normalised on a scale of 0 to 1; is a combination of the responses to the questions on why parents feel it is important to have children.
4. Normalised on a scale of 0 to 1; refers to child's own perception of well-being. This is a combination of the responses to the questions on how loved/comfortable child is at home.
5. Indicator of whether child belongs to a non-SC/ST and non-backward caste community.
6. An index (based on factor analysis) of the household's responses to questions of its kinship support base in community, how active it is (membership in societies etc.), level of trust within community.

Table 3: Summary Statistics: Indicators for Latent Variables, Age 1 and 5

Variable	Obs.	Median	Mean	Std. Dev.
Age 1				
Ante-natal Care[2]	1896	1.00	1.06	1.07
No Doc. at Birth	1923	0.00	0.47	0.50
Breastfeeding Duration[3]	1923	0.00	1.30	3.86
Unwanted Pregnancy	1910	0.00	0.08	0.27
Freq. See Dad[4]	1930	0.00	0.04	0.20
CG Depressed	1840	0.00	0.30	0.46
CG Education[4]	1950	15.00	11.75	4.41
Weight Z-Score	1934	-1.84	-1.835	0.937
Height Z-Score	1934	-1.63	-1.611	1.00
Age 5				
Peabody PVT Score	1950	34.05	44.22	29.03
CDA-Q Test Score	1950	10.00	9.51	2.68
Pre-School Status[5]	1950	0.00	1.74	1.83
CH Fluent in Native Lang	1950	1.00	0.90	0.30
CH Travels[6]	1950	3.00	2.46	0.87
Prop. Clothing Exp.	1925	0.50	0.62	0.34
Prop. Edu. Exp.	1925	0.20	0.24	0.25
Prop. Health Exp.	1950	0.13	0.13	0.17
Freq. See Dad	1913	2.00	1.91	0.31

Notes:
1. CH: Child; CG: Caregiver.
2. On a scale of 0-3; 0 indicates high level of AN care, 3 indicates no AN care.
3. Indicates number of months child was left without breastfeeding, from a recommended period of 16 months.
4. These variables have been rescaled; higher values indicate high neglect (higher freq of not seeing dad, and low level of caregiver education).
5. 0 indicates child doesn't attend pre-school; how the child fares in preschool (conditional on attendence) is given on a scale of 1-5, where 5 is excellent and 1 is poor.
6. Indicates whether child travels alone (1), with parents (2), with friends (3), or does not travel at all (0).

Table 4: Summary Statistics: Indicators for Latent Variables, Age 8 and 12

Variable	Obs.	Median	Mean	Std. Dev.
Age 8				
Raven's Score	989	23.00	22.97	5.30
Reading Level[2]	985	4.00	3.08	1.05
Writing Level[3]	964	2.00	2.09	0.68
Hyperactivity Score	993	2.00	1.73	0.59
Emotional Symptoms Score	993	2.00	1.34	0.86
Pro-Social Behaviour Score	994	2.00	1.70	0.64
Conduct Problems Score	994	2.00	1.24	0.91
Freq. See Dad[4]	965	2.00	1.87	0.41
Spend on Education	994	0.00	0.09	0.29
Does Child Work	994	1.00	0.67	0.47
Child Started School[5]	987	2.00	1.96	0.57
BMI Z-Score	994	-1.38	-1.39	0.993
Child Serious Illness	987	0.00	0.109	0.312
Age 12				
Peabody PVT Score	994	100.41	96.16	33.01
Math Test Score	994	6.00	5.91	2.64
Writing Level[3]	969	3.00	2.65	0.58
CH Friendliness Score	994	0.48	0.47	0.14
CH Group Membership	994	0.00	0.02	0.10
CH Self-Pride Score	994	0.64	0.64	0.13
CH Determination Score	994	0.72	0.71	0.12
CH Social Trust Score	994	0.95	0.88	0.20
Prop. Clothing Exp.	983	0.50	0.62	0.32
Prop. Edu. Exp.	983	0.30	0.31	0.22
Prop. Health. Exp.	994	0.13	0.13	0.17
Freq. See Dad[4]	913	2.00	1.84	0.40
BMI Z-Score	994	-1.51	-1.45	1.23
Child Serious Illness	987	0.00	0.24	0.43

Notes:
1. CH: Child.
2. On a scale of 1-4; 1 indicates cannot read at all, 4 indicates reads fluently.
3. On a scale of 1-3; 1 indicates cannot write at all, 4 indicates can write fluently.
4. Indicates whether child sees biological father, daily (2), monthly (1), once a year or never (0).
5. Indicates number of years since parents started school for child.

Table 5: Pairwise Spearman Rank Correlation Matrix

Table A: Age 1 and 5

	CS (Age 5)	NCS (Age 5)	Child Health (Age 1)	PI (Age 5)	PRF (Age 1)
Cognitive Skills (Age 5)	1.0000				
Non-cognitive Skills (Age 5)	0.8708**	1.0000			
Child Health (Age 1)	0.8688*	0.8617**	1.0000		
Parental Investment (Age 5)	0.7400**	0.7285**	0.7711**	1.0000	
PRF (Age 1)	-0.9341**	-0.8964**	-0.9308**	-0.7497**	1.0000

Table B: Age 8 and 12

	CS (Age 12)	NCS (Age 12)	CS (Age 8)	NCS (Age 8)	PI (Age 12)	PI (Age 8)
Cognitive Skills (Age 12)	1.0000					
Non-cognitive Skills (Age 12)	0.7641**	1.0000				
Cognitive Skills (Age 8)	0.9348**	0.6011**	1.0000			
Non-cognitive Skills (Age 8)	0.6231**	0.4276**	0.7346**	1.0000		
Parental Investment (Age 12)	0.7231**	0.7627**	0.7128**	0.7476**	1.0000	
Parental Investment (Age 8)	0.7914**	0.6373**	0.8586**	0.8328**	0.7607**	1.0000

Notes:

1. Standard Errors in parentheses.
2. + indicates significance at 10%; * at 5%; ** at 1%.

Table 6: Results 'Younger' Cohort

(a): Child Health - Age 1

Dependent variable:	Child Health (Age 1)
PRF (Age 0)	-0.414**
	(0.181)
Poor	-0.236**
	(0.072)
CH Male	0.211**
	(0.045)
CH Caste	0.279**
	(0.084)
CH Birth Order	-0.092**
	(0.028)
HH Size	0.004
	(0.016)
HH Urban	0.219**
	(0.097)
HH Non Ag. Occupation	0.001
	(0.029)
HH Mean Education	0.024**
	(0.010)
HH Social Connectedness	0.081
	(0.060)
Observations	**1950**

(b): Skill Accumulation - Age 5

Dependent variable:	Cognitive Skills (Age 5)	Non-cognitive Skills (Age 5)
Child Health (Age 1)	2.448**	-0.013
	(0.527)	(0.039)
Parental Investment	33.272**	1.854**
	(8.627)	(0.569)
Parent Altruism	25.686^{+}	-0.684^{+}
	(15.778)	(0.395)
CH Siblings	0.343	0.077*
	(0.793)	(0.049)
CG Educational Attainment	1.517**	0.047**
	(0.229)	(0.011)
HH Size	-0.359	0.013
	(0.432)	(0.022)
HH Social Connectedness	3.464**	-0.127^{+}
	(1.526)	(0.067)
HH Assets	1.860	1.112**
	(8.837)	(0.395)
Observations	**1950**	**1950**

Notes:
1. Standard Errors in parentheses.
2. $^{+}$ indicates significance at 10%; * at 5%; ** at 1%.
3. CH: Child; CG: Caregiver; HH: Household.

Table 7: Results 'Older' Cohort

(a): Skill Accumulation - Age 8

Dependent variable:	Cognitive Skills (Age 8)	Non-cognitive Skills (Age 8)
Parental Investment	10.877**	0.642
	(3.406)	(0.392)
CH Health	0.551*	0.023
	(0.262)	(0.059)
CH Well-being	0.695	-0.093
	(0.701)	(0.144)
CH Siblings	0.297**	0.001
	(0.128)	(0.030)
CH Male	0.513	0.056
	(0.440)	(0.060)
CH Caste	-0.387	0.086
	(0.312)	(0.087)
Public School	-1.455*	-0.005
	(0.630)	(0.088)
CG Education	-0.059	0.009
	(0.062)	(0.009)
HH Size	-0.142+	-0.027+
	(0.079)	(0.014)
HH Social Connectedness	0.672+	0.054
	(0.349)	(0.087)
HH Assets	0.284	-0.078
	(0.238)	(0.045)
HH Urban	-0.560**	0.282*
	(0.797)	(0.115)
HH Wage	0.251*	0.042
	(0.121)	(0.030)
HH Mean Education	-0.007	-0.007
	(0.039)	(0.008)
Observations	**994**	**994**

Notes:
1. Standard Errors in parentheses.
2. + indicates significance at 10%; * at 5%; ** at 1%.
3. CH: Child; CG: Caregiver; HH: Household.

Table 8: Results 'Older' Cohort

(b): Skill Accumulation - Age 12

Dependent variable:	Cognitive Skills (Age 12)	Non-cognitive Skills (Age 12)
Cognitive Skills (Age 8)	3.900**	0.011**
	(1.509)	(0.004)
Non-cognitive Skills (Age 8)	-0.534	-0.001
	(1.045)	(0.005)
Parental Investment	9.989**	0.196**
	(3.823)	(0.046)
CH Health	0.282	-0.003
	(0.253)	(0.002)
Parent Altruism	0.014	-0.013
	(3.108)	(0.017)
CH Well-being	1.354	0.077**
	(1.998)	(0.029)
CH Caste	-1.237	0.003
	(0.652)	(0.004)
CH No. of Years of Schooling	-0.365	0.002
	(0.437)	(0.003)
Public School	3.045**	0.061**
	(1.137)	(0.013)
CG Educational Attainment	0.004	0.001
	(0.094)	(0.001)
HH Size	-0.033	0.000
	(0.138)	(0.001)
HH Social Connectedness	0.751+	0.002
	(0.443)	(0.003)
HH Assets	2.238+	0.047**
	(1.903)	(0.011)
HH Urban	0.827	0.004
	(1.219)	(0.008)
HH Wage Recipient	-1.056*	-0.008**
	(0.474)	(0.003)
HH Mean Education	0.111+	0.000
	(0.060)	(0.001)
Observations	**994**	**994**

Notes:
1. Standard Errors in parentheses.
2. + indicates significance at 10%; * at 5%; ** at 1%.
3. CH: Child; CG: Caregiver; HH: Household.

APPENDIX: Factor Models

Table 9: Factor Loading: Explanatory Variables; 'Younger' Cohort

Dependent variable:	Parental Investment (Age 5)	Psycho Risk Factors (Age 0)
Unwanted Pregnancy	- -	0.082** (0.026)
No. Doc. at Birth	- -	0.729** (0.083)
Breastfeeding Duration	- -	-2.66** (0.436)
Ante-natal Care	- -	1.00
CG Education	- -	16.190** (1.582)
CG Depression	- -	0.344** (0.054)
Freq. See Dad	- -	-0.006 (0.018)
Prop. Clothing Exp.	1.00	- -
Prop. Edu. Exp.	1.209** (0.076)	- -
Prop. Health Exp.	0.105** (0.027)	- -
Freq. See Dad	-0.092* (0.052)	- -

Table 10: Factor Loading: Outcome Variables; 'Older' Cohort

Dependent variable:	Cognitive Skills (Age 12)	Non-cognitive Skills (Age 12)	Cognitive Skills (Age 8)	Non-cognitive Skills (Age 8)
Peabody PVT Score	1.00	- -	- -	- -
Math Test Score	0.282** (0.071)	- -	- -	- -
Writing Level	0.059** (0.014)	- -	- -	- -
CH Friendliness Score	- -	1.00	- -	- -
CH Group Membership	- -	0.133** (0.053)	- -	- -
CH Self-Pride Score	- -	1.3** (0.178)	- -	- -
CH Determination Score	- -	1.078** (0.17)	- -	- -
CH Social Trust Score	- -	2.728** (0.486)	- -	- -
Raven's Score	- -	- -	1.00	- -
Reading Level	- -	- -	0.543** (0.144)	- -
Writing Level	- -	- -	0.103** (0.045)	- -
Emotional Symptoms Score	- -	- -	- -	1.00
Conduct Problems Score	- -	- -	- -	0.518** (0.134)
Hyperactivity Score	- -	- -	- -	0.165** (0.083)
Pro-Social Behaviour Score	- -	- -	- -	0.253** (0.078)

Table 11: Factor Loading: Explanatory Variables; 'Older' Cohort

Dependent variable:	Parental Investment (Age 12)	Parental Investment (Age 8)
Prop. Clothing Exp.	1.00	- -
Prop. Edu. Exp.	2.76** (0.517)	- -
Prop. Health. Exp.	0.098** (0.039)	- -
Freq. See Dad	0.314 (0.185)	- -
Spend on Education	- -	1.00
Freq. See Dad	- -	-0.323 (0.493)
Child Started School	- -	-21.102
Does Child Work	- -	-1.189** (0.48)

Table 12: Factor Loading: Outcome Variables; 'Younger' Cohort

Dependent variable:	Cognitive Skills (Age 5)	Non-cognitive Skills (Age 5)	Child Health (Age 1)
CDA-Q Test Score	0.068** (0.005)	- -	- -
Peabody PVT Score	1.00	- -	- -
Pre-School Status	- -	1.00	- -
CH Travels	- -	-0.345** (0.039)	- -
CH Fluent in Native Lang.	- -	-0.007 (0.011)	- -
Weight Z-Score	- -	- -	1.00
Height Z-Score	- -	- -	1.187** (0.085)

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 289966
Email: younglives@younglives.org.uk

www.younglives.org.uk