CSAE WPS/2010-14

Health, Nutrition and Academic Achievement: New Evidence from India*

Geeta Kingdon†
Institute of Education
University of London

Courtney Monk
Centre for the Study of African Economies
University of Oxford

June 2010

Abstract

Using new and unique panel data, we investigate the role of long-term health and childhood malnutrition in schooling outcomes for children in rural India, many of whom lack basic numeracy and literacy skills. Using data on students' performance on mathematics and Hindi tests, we examine the role of the endogeneity of health caused by omitted variables bias and measurement error and correct for these problems using a household fixed effects estimator on a sub-sample of siblings observed in the data. We also present several extensions and robustness checks using instrumental variables and alternative estimators. We find evidence of a positive causal effect of long-term health measured as height-for-age z-score (HAZ) on test scores, and the results are consistent across several different specifications. The results imply that improving childhood nutrition will have benefits that extend beyond health into education.

Keywords: Health, Nutrition, Schooling, India
JEL Classification Codes: I12, I21

*The data used in this paper were collected by Geeta Kingdon at the Centre for the Study of African Economies (CSAE), Oxford, in collaboration with Rukmini Banerji of Pratham in 2007-2008. The survey was funded by the Spencer Foundation. The George Webb Medley Fund also provided funding for one of the author's participation in the pilot. The authors would like to thank Francis Teal, Stefan Dercon, Paul Glewwe, and William Masters for helpful comments, and to recognize Rukmini Banerji for her help on the tests and questionnaires. Any omissions, errors and opinions expressed herein are the sole responsibility of the authors.

†Corresponding author: G.Kingdon@ioe.ac.uk; Department of Education and International Development, Institute of Education, University of London, 20 Bedford Way, London, WC1H 0AL, United Kingdom.

1 Introduction

India certainly has good reason to be concerned about education. Although India is somewhat comparable to its neighbours Bangladesh and Pakistan, it lags far behind China and other developing countries in Asia and South America (UNESCO, 2006). Despite progress in terms of primary school enrollment, attendance and retention rates are low, achievement levels are unimpressive, and teacher absenteeism is high (Kingdon, 2007; Kremer, Chaudhury, Rogers, Muralidharan, & Hammer, 2005). The 2007 Annual Status of Education Report (ASER) shows some alarming statistics for rural Indian children: "Nearly 40% of children at class 5 level cannot read a class 2 text, while around 60% at that level are unable to carry out simple divisions. What should be more worrying though, is the fact that in class 2, only 9% of children can read the text appropriate to them, and 60% cannot even recognize numbers between 10 and 99" (Pratham, 2008).

While there is much research seeking to discover the determinants of student outcomes, this has focused mostly on factors thought to be most amenable to policy influence, namely school and teacher factors (see Glewwe (2002) and ? (?) for a review). However, child and home background factors are much more versatile indicators of the child's circumstances and may be the more important factors behind children's educational outcomes.

In this paper, we focus on health as a determinant of primary school outcomes. Health is in general an under-studied component of educational outcomes, particularly in developing countries. However, there is good reason to believe that health, and specifically malnutrition, is a key part of child development. Indeed, research has found that malnutrition in the early years of life may cause later deficiencies in cognitive development. There is some disagreement as to whether the effects of malnutrition on cognitive development are most acute when the child is a baby (0-6 months, as Dobbing (1976) argues) or a toddler (6-36 months), but there seems to be a consensus that the first two years are very important (Waber et al., 1981; Pollitt, Gorman, Engle, Rivera, & Martorell, 1995; Martorell, 1995; Glewwe & King, 2001). In the long run, if unhealthy children do poorly in school, and low schooling attainment compromises labour market performance, then malnutrition in childhood can undermine labour outcomes in adulthood (Deaton, 2008)—this effect will be compounded if unhealthy children are more likely to become unhealthy, uneducated adults.

As with education, Indian children also suffer from below average health. Even though India experienced improvements in overall nutrition in the 1990s, the advances in anthropometric measures are "slow relative to what might be expected... in light of India's recent high rates of economic growth" (Deaton & Drèze, 2008, p.1). Statistics from the 2005-06 wave of the National Family Health Survey also show only slow improvements in childhood malnutrition: from 1998-99 to 2005-06, the proportion of undernourished (based on weight-for-age) children was essentially unchanged (47% to 46%), and the percentage of stunted (based on height-for-age) only declined about 1 percentage point per year from 45% to 38% (Citizens Initiative for the Rights of Children Under Six, 2006).[1] This is especially true for girls, and for children in rural areas (Tarozzi & Mahajan, 2007; Tarozzi, 2008). Compared to Western standards, India still has a long way to go (see Table 1).

Following on from the research by nutritionists, it then becomes important to find out how nutrition and health later affect academic outcomes once children enter school. And in developing countries like India, where child malnutrition is common, understanding the impact of health on schooling is of even greater importance and urgency.

One might expect intuitively that unhealthy children do worse in school on average, for several reasons: their cognitive development has been compromised, they are more likely to be absent from school, they have less time or energy to finish homework, they have problems concentrating, etc. On the other hand, sickly children may do less outside work or play, staying home and studying more than their healthier counterparts—this effect might be less prevalent than the first but nonetheless plausible. Previous research that has investigated the role of

[1] The FOCUS report (Citizens Initiative for the Rights of Children Under Six, 2006) does not state if these differences are statistically significant.

Table 1: UNICEF Nutrition Statistics, 2000-2006

	India	China	Bangladesh	Pakistan	Vietnam	US
Infants with low birthweight[a]	30%	2%	22%	19%	7%	8%
Children exclusively breastfed (< 6 mos)[a]	46	51	37	16	17	–
Children breastfed with solid food (6-9 mos)[a]	56	32	52	31	7	–
Under-5s underweight, moderate & severe	43	7	48	38	25	2
Under-5s underweight, severe	16	–	13	13	3	0
Under-5s wasting, moderate & severe	20	–	13	13	7	0
Under-5s stunting, moderate & severe	48	11	43	37	30	1

Source: UNICEF, http://www.unicef.org/infobycountry, accessed 6 November 2008.
[a]: Figures for 1999-2006.
Definitions: Low birthweight - Less than 2,500 grams. Underweight - Moderate and severe - below -2 standard deviations from median weight for age of reference population; severe - below -3 standard deviations from median weight for age of reference population. Wasting - Moderate and severe - below -2 standard deviations from median weight for height of reference population. Stunting - Moderate and severe - below -2 standard deviations from median height for age of reference population.

health in academic performance has struggled to find convincing evidence of a direct, causal link between health and school achievement.[2] This is in part because much of the research to date has relied almost exclusively on cross-sectional data, and in part because most surveys lack comprehensive information on both the student's background and his/her school (see, e.g. Gomes-Neto, Hanushek, Leite, and Frota-Bezzera (1997)). Both of these factors mean that endogeneity caused by unobserved omitted variables is a real concern that has not always been adequately addressed.

The question of causality is therefore still open-ended, although there has been some good progress in the last decade or so. Not only has there been some experimentation (for instance Bobonis, Miguel, and Puri-Sharma (2006); Miguel and Kremer (2004); Tan, Lane, and Lassibille (1999); Pollitt et al. (1995); Martorell (1995)), but the approaches using nonexperimental data have also improved. Beginning with Glewwe and Jacoby (1995), there has been a clear focus in the literature on using fixed effects and instrumental variables (IV) to tease out the causal relationship between long-term health (measured as height-for-age) and educational outcomes. Using cross-sectional data from Ghana, Glewwe and Jacoby (1995) instrument for height-for-age (HAZ) using community health variables, school characteristics, and parental and household characteristics in a model of school enrollment. They find that child malnutrition is a significant determinant of delayed enrollment and explains the delay more than other possible factors. Also using Ghanaian data, Behrman and Lavy (1997) examine the effect of HAZ on cognitive

[2]Because height-for-age (HAZ) is one of the best indicators of long-run nutritional status (Glewwe & King, 2001), we use long-term health, HAZ and nutrition as roughly interchangeable concepts.

achievement tests, focusing on the possible biases induced by omitted variables and measurement error. They note that the direction of the bias is unknown *a priori* and then compare IV results with community and family fixed effects to see what we can learn about the unobserved variables. They conclude that using family and community level instruments imply a downward bias to the OLS results, but that the family and community fixed effects suggest an upward bias. Their preferred estimates are the fixed effects results, as they claim that the assumption that there are no unobserved educational inputs in the error term biases the results upwards even if one has a valid instrument. They conclude that child health does not affect cognitive achievement. A recent working paper by Wisniewski (2009) uses price and weather shocks as instruments and does find that both nutritional status and other health problems like intestinal worms have a significant impact on tests scores in the OLS, IV and school fixed effects results.

Longitudinal data have been in general scarce, but there are a few important exceptions. Glewwe, Jacoby, and King (2001) use a siblings fixed effects approach with Filipino data. They difference HAZ across siblings, and then instrument this difference with an earlier observation of sibling height. Their results show a huge downward bias in the OLS—the 2SLS coefficient on HAZ is about 30 times the OLS coefficient—which they attribute mostly to measurement error. The authors argue that better nourished children perform better in school because of both early enrollment and better productivity once enrolled. Another recent study is Alderman, Behrman, Lavy, and Menon (2001), which uses panel data of children in rural Pakistan. Based on a structural model (that differs from the one used in this paper), the authors argue that food prices in previous periods are valid instruments for child health in those periods. They find that better health raises the probability of enrollment in school at age 7. Glewwe (2005) points out that Alderman et al. (2001) may overestimate these health effects because past food prices could affect savings for education in those periods, which would then violate the required assumption that the instruments have no effect on schooling other than an indirect effect through past health. Finally, Alderman, Hoddinott, and Kinsey (2006) use a maternal fixed effects (essentially a siblngs fixed effects as in Glewwe et al. (2001)) estimation with Zimbabwean longitudinal data to measure the long-term effects of malnutrition. With war and drought as instruments for early childhood HAZ, they find that higher HAZ at pre-school age increases height as a young adult, raises completed years of schooling, and lowers the age of enrollment. Though their instruments are potentially weak and their sample size is small, their results are nonetheless compelling.

This paper improves on existing research by using new panel data from rural India, which allow us to investigate the role of child health and nutrition in academic achievement as measured by test scores. These unique data permit several levels of fixed effects analysis, including child fixed effects. We believe that we are able to identify the causal impact of early childhood nutrition and health, as measured by HAZ, on educational outcomes.

In the next section we discuss a theoretical and empirical model linking nutrition and academic tests scores, and Section 3 explains the data in more detail. Empirical results are presented in Section 4, in which we pare down the sample to analyze siblings and introduce household fixed effects. Robustness checks, including instrumental variables analyses and fixed effects at the child level, are presented in Section 5. A final section concludes.

2 Theoretical model

There are different ways to model childhood nutrition and schooling production, and the method selected will have implications for our estimation strategy, and in particular, our instrument set. We adopt here a simple framework developed by and attributable to Glewwe and Miguel (2008) that relates child health and nutrition to academic outcomes. We will first develop a structural production function, and then derive conditional demand functions that we then estimate with our data.

The first assumption underpinning the model is that there are two time periods: (1) birth to 6 years when the child enters school; and (2) school age, from about age 6 to puberty. These periods mirror the findings of nutritionists, who focus on period 1 (specifically birth to 24 months) for cognitive development. With these periods defined, we can write down a *structural* production function for schooling outcomes—in this case test scores (T_2)—observed in period 2:

$$T_2 = T_{2,P}(H_1, H_2, PEI_1, PEI_2, S_2, \alpha, Q, YS_2) \tag{1}$$

where the subscript "P" indicates that this is a production function. H_t measures nutritional health from $t-1$ to t, PEI_t is parental education inputs in period t (such as time inputs and school supplies), S_2 is any individual shock to schooling performance experienced in period 2 (such as temporary illness[3] or absences from school), α denotes the child's innate schooling ability, Q is school and teacher characteristics and quality (assumed to be time invariant), and YS_3 is years of schooling completed by period 2 (which is highly dependent on the child's age). This production function is modified from Glewwe and Miguel (2008) to include contemporary shocks in period 2 that could affect test performance.

Parents maximize a utility function:

$$U = U(H_1, H_2, C_1, C_2, T_2) \tag{2}$$

where C_t is parental consumption of an aggregate consumption good in period t. There are four constraints in the parents' optimization problem. The first constraint is the production function in Equation 1. Two other constraints are the production functions for child health:

$$H_1 = H_1(C_1^c, M_1, HE_1, \eta) \tag{3}$$

$$H_2 = H_2(C_1^c, C_2^c, M_2, HE_2, \eta) \tag{4}$$

where C_1^c is the child's consumption of the aggregate consumption good in period t; M_t is health care inputs in period t; HE_t is the local health environment, e.g. water and sanitation quality, in period t; and η is the innate healthiness of the child. HE_t and η are considered exogenous to the parents' control. The final constraint is an intertemporal budget constraint:

$$W_0 = p_{C,1}(C_1 + C_1^c) + \frac{p_{C,2}(C_2 + C_2^c)}{(1+r)} + p_{M,1}M_1 + p_{PEI}PEI_1 + \frac{p_{M,2}M_2 + p_{PEI}PEI_2 + p_S YS_2}{1+r} \tag{5}$$

[3]Temporary illness would not be captured in the HAZ variable H_2, as this is a long-term measure of health.

W_0 is initial wealth, and r is the interest rate. $p_{\cdot,t}$ are the prices of the various types of goods in period t, but p_{PEI} is assumed to be constant over the two periods, and p_S is the price of schooling in period 2.

Optimizing Equation 2 with respect to Equations 1, 3, 4 and 5 yields the following demand functions:

$$\begin{aligned}
C_t &= C_{t,D}(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{PEI}, p_S; \\
&\quad HE_1, HE_2, Q, ME, FE; \alpha, \eta, \sigma, \tau) \quad t = 1,\, 2 && (6) \\
C_t^c &= C_{t,D}^c(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{PEI}, p_S; \\
&\quad HE_1, HE_2, Q, ME, FE; \alpha, \eta, \sigma, \tau) \quad t = 1,\, 2 && (7) \\
M_t &= M_{t,D}(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{PEI}, p_S; \\
&\quad HE_1, HE_2, Q, ME, FE; \alpha, \eta, \sigma, \tau) \quad t = 1,\, 2 && (8) \\
PEI_t &= PEI_{t,D}(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{PEI}, p_S; \\
&\quad HE_1, HE_2, Q, ME, FE; \alpha, \eta, \sigma, \tau) \quad t = 1,\, 2 && (9) \\
YS &= YS(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{PEI}, p_S; \\
&\quad HE_1, HE_2, Q, ME, FE; \alpha, \eta, \sigma, \tau) \quad t = 1,\, 2 && (10)
\end{aligned}$$

where the subscript "D" indicates that these are demand functions. ME and FE are mother's and father's education, respectively; σ is parental tastes for schooling, and τ is parental tastes for child health.

We can then construct a *conditional demand* equation for academic achievement. To do so, we first assume that given the utility maximizing levels of C_1^c, C_2^c, M_1 and M_2, which yields conditional demand equations for parental inputs and years of schooling:

$$\begin{aligned}
PEI_t &= PEI_{t,D}(H_1, H_2; W_{CD}, \omega) \quad t = 1,\, 2 \\
YS &= YS(H_1, H_2; W_{CD}, \omega) && (11)
\end{aligned}$$

where W_{CD} is the household's non-health expenditures and equals $W_0 - p_{C,1}C_1^c - p_{M,1}M_1 - (p_{C,2}C_2^c - p_{M,2}M_2)/(1 + r)$ and ω is the vector $(r, p_{C,1}, p_{C,2}, p_{PEI}, p_S; Q, ME, FE; \eta, \sigma, \tau)$. One can then substitute these conditional demand functions into the production function in Equation 1, yielding:

$$T_2 = T_{2,CD}(H_1, H_2; S_2, W_{CD}, \omega, \alpha) \tag{12}$$

This is a conditional demand function because it is not a structural equation like Equation 1, but it is not a reduced form because H_2 is endogenous (Glewwe & Miguel, 2008). Note that H_1 is predetermined at the start of period 2, as parents make their decisions at the start if the period. This equation shows the net effect of H_2, i.e. both the direct effect and the indirect effect(s) of H_2 on T_2. Indirect effects are those such as a change in parental inputs or years of schooling as a result of a change in health. In other words, this equation can show how small changes in H_1 or H_2 can affect T_2, allowing for parents to alter their input choices in response to these changes. Since we are not prepared to make a strong separability assumption, we do not presume that the effect of H_2 in Equations 1 and 12 are the same; however, we do believe that it is worth estimating the conditional demand function to find out the relationship between health and academic achievement. Moreover, the estimation is useful to policymakers, who can themselves change the price of schooling, the price of health inputs, and the health environment through policy efforts. *A priori*, the signs of H_1 and H_2 in Equation 12 should be positive (Glewwe & Miguel, 2008).

2.1 Causality

The causal effects of H_1 and H_2 in Equation 12, however, are not easy to estimate. A central issue is that some of the variation in anthropometric indicators is perfectly natural (some children are naturally tall, others are short) and has nothing to do with their nutritional intake, and likewise should not affect their test scores. In addition, measurement error is a confounding issue. Another problem is endogeneity caused by the fact that decisions about schooling, health and nutrition are not randomly determined but are rather based at least to some extent on household decisions (Behrman, 1996). Thus associations that are commonly observed between health and schooling may not in fact reflect true *causality*, resulting in omitted variable bias

Table 2: Variables Influencing Cognitive Achievement

Variables	Observed	Unobserved
Individual	Anthropometric indicators	Innate ability
	Gender	Motivation
	Age	Genetic Endowment
	Illness	Capacity to concentrate
	Attendance & enrollment	
Family	Parental education	Household intellectual atmosphere
	Parental occupation	Parental time and involvement in child's development
	Assets	Income
		Nutritional inputs (food, vitamins, etc.)
		Parental preferences for child health
Community	Availability of health services	General intellectual atmosphere
	Prevalence of disease	Water quality and sanitation
	Education resources	Local labour market conditions
School	Facilities	Effectiveness of school management
	Public/private	
Teacher	Qualifications/certifications	Innate teaching ability
	Experience	Teacher effort

Source: Author's interpretation of the School-Tells data. This table is inspired by Table 1 in Behrman (1996, p.25).

with the OLS results. Table 2 has a list of many variables observed and unobserved in the data; though it is not an exhaustive list (see Table 13 in Appendix B for all of the variables used), it does give a good sense of the scope of the data.

An example of an omitted variable that could cause bias is an unobservable parental characteristic (such as "good parenting skills") that is positively correlated with both health and education that creates the appearance of a positive causal effect that is in fact a simple upward bias in the OLS. On the other hand, there may be a downward bias if healthy children are actually more likely to spend time doing things other than schoolwork because the returns to education are sufficiently low in their community. Moreover, measurement error is likely to be a problem in any dataset with anthropometric measurements of children, especially those collected in the field. In addition, measuring age accurately in a setting such as rural India is problematic.[4] Furthermore, classical measurement error causes attenuation bias and an underestimate of the effect of health. In sum, the overall sign of the bias is ambiguous. These endogeneity concerns are particularly acute, even if we have a very comprehensive set of information on households and schools.

[4] We did make a concerted effort to 'clean' our height and age data as much as possible to avoid egregious measurement errors (see Section A.1).

In this paper, we focus primarily on the relationship between children's health H_1 and their test scores. We do not have a perfect measure of H_1 in the data, since the children are observed twice at school age, i.e. twice during period 2. However, we argue here that HAZ is a reasonably good indicator of health status achieved by the end of period 1, H_1, because it is mostly the product of the nutritional inputs a child has experienced during the first period. By the same reasoning, HAZ is not a good measure of H_2. Indeed, HAZ does not reflect short-run changes to health in period 2 because it does not fluctuate as a short-term measure should. In fact, Tanner, Healy, Lockhart, MacKenzie, and Whitehouse (1956) find that the correlation between height at 2 years of age and adult height is on the order of 0.7 to 0.8, depending on gender; this implies that children's height at older ages, especially before puberty, mostly reflects nutritional and medical inputs experienced when very young and not current inputs. Therefore, though measured with error, we contend that height-for-age observed in our dataset is a reasonable measure of H_1.[5]

Following Glewwe et al. (2001) and citing the nutritional literature, we then make the strong assumption that the effect of H_2 is negligible in Equations 1 and 12 (this assumption is later relaxed in Section 5). That is, the strongest effects of child health on schooling outcomes come early in period 1, in the first two years of life (Waber et al., 1981; Pollitt et al., 1995; Martorell, 1995; Glewwe & King, 2001). As such, HAZ contains the most information about malnutrition experienced in early childhood—which is the most important for cognitive development, and, by extension, for schooling performance. A history of chronic malnutrition prevalent in developing countries, and especially in rural India, will also show up most readily in long-term as opposed to short-term health.[6]

Assuming a linear form for Equation 12, we can expand the model:

$$T_2 = \beta_0 + \beta_1 H_1 + \beta_2 S_2 + \beta_3 W_{CD} + \beta_4 r + \beta_5 p_{C,1} + \beta_6 p_{C,2} + \beta_7 p_{PEI} + \beta_8 p_S + \beta_9 Q +$$

[5] An important limitation of this approach and of our data is that we are not able to examine the effect of malnutrition at different points in childhood, as do Glewwe and King (2001).

[6] Some research has been done on academic achievement and obesity in developed countries (Kaestner & Grossman, 2008; Averett & Stifel, 2007; Sabia, 2007; Datar & Sturm, 2006). Obesity is uncommon among rural Indian children and is therefore not explored in this paper.

$$\beta_{10}ME + \beta_{11}FE + \beta_{12}\eta + \beta_{13}\sigma + \beta_{14}\tau + \beta_{15}\alpha \tag{13}$$

Building further on Equation 13, we can begin to address the endogeneity issues discussed above. After presenting simple OLS results for the full sample for completeness, we narrow down our sample to a sample of siblings and conduct family fixed effects on this sub-sample. This approach is helpful because the unobserved variables that are constant at the level of the household will be differenced out in a household fixed effects analysis of a sub-sample of siblings. For example, to the extent that parental tastes for their children's health τ is constant across their children, such tastes will be removed by the differencing. The household fixed effects significantly cuts down on the sources of potential biases. Unfortunately, the sample was not collected as a sample of siblings; rather, the sub-sample of siblings is comprised of brothers and sisters that we just happen to choose as part of the greater sampling process. This means that the inference with the siblings sub-sample may be weaker due to the smaller size of the sample.

As a further robustness check, we pursue an instrumental variables strategy to address remaining endogeneity problems with HAZ. It is difficult to find a valid instrument for H_1, as it must be correlated with H_1 but not with the unobserved variables and measurement errors. In addition, within a siblings-only framework, it must vary across siblings within the household. Possible candidates are period 1 health prices. However, one can criticize the period 1 instruments for being unlikely to change and therefore related to health and prices in period 2. If true, then these period 1 health prices may be invalid, given the specification in Equation 12.

An additional alternative, and one that is more likely to be available, is to use an exogenous shock that affected H_1 but is unrelated to period 2 academic performance. One such instrument is weather data (Wisniewski, 2009). Although flood data are not available for India, rainfall data are available for the period 1990-2002 at the district level. Therefore we use rainfall data from the child's birth year, year one and year two of life to instrument for H_1, as these should correspond to the years that matter most for cognitive development. However, it is important to keep in mind that weather data are likely to provide only weak instruments, and the IV results are thus presented as a robustness check for the household fixed effects. Furthermore, if weather shocks in period 1 affected food prices in period 1, and we do not control for initial wealth excluding spending on health inputs in period 1 (which we cannot do because we do not observe period 1 spending), then rainfall may violate the exclusion restriction that it only affect T_2 through H_1 and not through any other unobserved variable (Glewwe & Miguel, 2008).

A final robustness check that we undertake is to use fixed effects at the child level. This approach should eliminate any child-level characteristics that remain constant, such as inherent ability, therefore providing another avenue through which to bolster our results. However, child fixed effects cannot eliminate all sources of endogeneity that vary over time. In addition, the type of instruments that would be necessary for such a specification—ones that vary over time within the same student—are really difficult to find and then justify.

3 Data

The School-Tells study was designed to learn about schools, teachers, and overall achievement in some of the most educationally disadvantaged parts of rural India. We first ran a pilot study in March 2007 in Bihar to vet the testing tools and the questionnaires. Building on what we learned in the pilot, we then ran the full survey in the field from July 2007 to April 2008 (corresponding to the north Indian academic year). The survey was administered in 11 districts, five in Uttar Pradesh (UP) and six in Bihar (see Appendix Section A for more details on sampling).[7]

The study consisted of several questionnaires designed to measure all aspects of the children's learning environment. Altogether, we collected information at the child, household, teacher, school, and village levels. The enumerators visited the schools four times, at roughly equal intervals, over the nine months of the school year 2007-08, and collected school level information

[7]Districts are: Agra, Bijnor, Lucknow, Mahoba, Shrawasti, Banka, Bhojpur, Gaya, Madhepura, Madhubani, and Saran.

during each of these visits. The testing of the children occurred in the first visit and the last visit. Anthropometric data and short-term health measures were collected on the same day as the test, at the start and at the end of the school year (nine months apart). Detailed household, village and teacher questionnaires were also administered, but most of the information on these was collected only on the first visit to each school.

3.1 Anthropometric indicators

Although health is a function of many things that can vary from child to child, particularly genetics, anthropometric indicators are a good measure of overall physical development, as they reflect diet and growth experienced in the child's lifetime. To discover how "normal" a child's growth is, we can compare him/her to a "healthy" reference group (say, children from the United States) (O'Donnell, Doorslaer, Wagstaff, & Lindelow, 2007). Thus all of the anthropometric indicators are transformed into z-scores that map our Indian children to a reference population of US children.[8]

Table 3: Anthropometric Statistics, Pooled Waves – All Children in Classes 2 and 4

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
HAZ	-1.27	1.37	-5.999	4.618	7,924
BAZ	-1.50	1.05	-4.995	1.633	7,625
WAZ	-1.85	1.27	-5.976	2.458	7,705

Source: Author's calculations based on the pooled observations across both waves. Summary statistics reflect cut-offs described in Section 3.1.

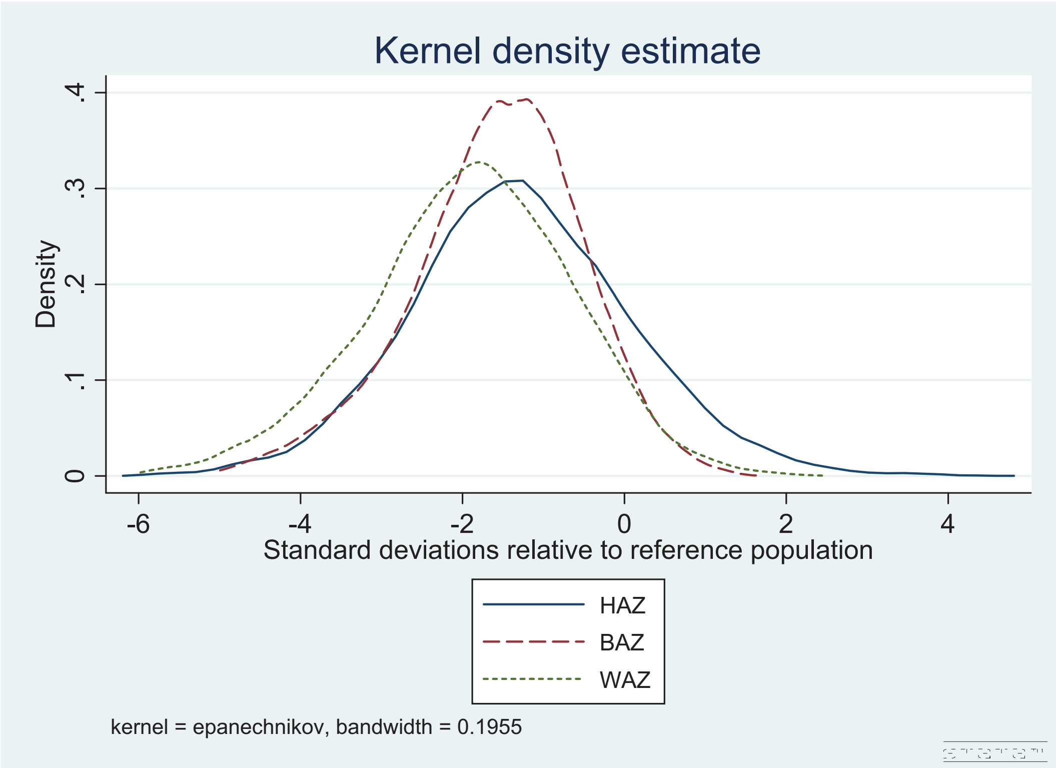
Following the existing literature, the main indicator that we focus on is height-for-age (HAZ), which, for children, is a cumulative (linear) measure of lifetime nutritional and medical inputs.[9] It is a long-term measure and cannot capture short-term fluctuations in malnutrition. Severely low HAZ is referred to as "stunting." In addition, we have data on BMI-for-age (BAZ), which is a medium- to short-term indicator of malnourishment than HAZ, and weight-for-age (WAZ).[10] In a developing country, HAZ can be as low as -6 and as high as 5. To put the scale of HAZ into perspective, a 1 unit change in HAZ (i.e. 1 standard deviation of the reference

[8] We use the 2000 CDC growth charts. See Appendix A.1 for more details.

[9] A chronic illness that has as a side effect malnutrition can adversely affect child growth as well.

[10] Very low WAZ is considered "underweight." WAZ is a composite measure of weight-for-height (WHZ) and HAZ.

Figure 1: Distribution of Anthropometric Indicators, Pooled Waves – All Children in Classes 2 and 4



distribution) is about 6cm of raw height in our data. The average growth for the children was 4.6cm over the nine months, one unit of HAZ is roughly equivalent to a year's worth of height gain. Summary statistics of the anthropometric indicators can be found in Table 3, and their Kernel densities (empirical distributions) are graphed in Figure 1. They are roughly normally distributed, though with a mean below zero because their health is below the reference population's, indicating that the children's health is sub-par. The anthropometric indicators required some cleaning, discussed in Appendix A.1.[11]

3.2 The tests

We crafted the tests with help from Pratham, an Indian NGO, which is also responsible for designing the testing tools for the annual ASER survey of learning achievement levels. The tests

[11] Each anthropometric indicator for the reference population has a standard normal distribution with mean zero and standard deviation of one. Our sample has distributions of HAZ, BAZ and WAZ that are shifted to the left, i.e. the means are below zero.

assessed the full range of academic ability: reading letters, words, sentences, and paragraphs; writing letters, words, and sentences; recognizing 1- and 2-digit numbers; and solving numeric and word math problems.[12] Children in both classes 2 and 4 were given the same exact test, at both the start and end of the school year. The test covered relevant material from the standard class 2 and class 4 textbooks and matches very closely material in existing textbooks. The test consisted of four sections: writing, reading, simple mathematics (including numeracy) and advanced mathematics. None of the material was meant to be beyond a class 4 level. The subject tests have a mathematics score out of 173 possible points and a Hindi score out of 241 possible points.

As the 2007 ASER report shows, many children in rural India struggle with the most basic academic tasks. Although there are some who manage to thrive even in home and school environments that are hardly conducive to learning, most children are not able to do work at their grade level. Some children do achieve and even surpass the expected level; conversely many others, despite attending school fairly regularly, do not learn to read or write. It is not uncommon, especially in some states including Uttar Pradesh and Bihar, to find children who do not possess any literacy and/or numeracy skills. Indeed, in several districts across these two states, more than 40% of children in classes 1 and 2 cannot read letters of the alphabet (Pratham, 2008).

The testing tools thus begin with very simple questions, such as identifying letters and numbers. The questions increase in difficulty gradually until reaching a class 4 standard. Therefore, by design, part of the test is too difficult for a class 2 student, and thus we expect the class 2 scores to be automatically below the class 4 marks. Nonetheless both the Hindi and math tests do have several questions that class 2 students should know how to answer, particularly the ones that test recognition of letters and single digit numbers.

[12] The testing materials are available upon request.

3.3 Dependent variable

To characterize the dependent variable for our analysis, we normalize the total score, setting the mean to zero and the standard deviation to one. To do so, we pool the scores by subject across both waves and normalize that distribution (we can do this because we administer the same test in both time periods).[13] This permits us to see how a child's second test score changes relative to his/her own performance on the previous test. The normalization also allows for more direct comparability with the results of other studies that also normalize achievement test scores.

3.4 The sample

We restrict the sample to children whom we observe twice and who are between 6 and 14 years old, inclusive. With the combination of chaotic classroom assignments, multi-grade grouped classes, and delayed enrollment in rural Indian schools, we felt that these were reasonable cut-offs for children who should be in second and fourth grades.[14] We drop children who are missing HAZ measurements at two periods in time and focus on the panel individuals. We include all children who meet these criteria, regardless of whether their scores went up or down over time.[15] This exercise results in a "pooled sample" of 3,921 children who are observed twice with non-missing values for relevant variables. This sample is used only in Section 4.2. The more important "sibling sub-sample", used in subsequent sections of the paper, is comprised of 718 siblings in 356 households. Altogether, the sibling sub-sample contains 1,436 observations (two per sibling).

[13] When conducting the analysis by class and by gender, we re-normalize the subject scores.

[14] For details on the age distribution in primary school grades, see Pratham (2009).

[15] There are a few children in the sample whose scores increased considerably—for example by 300 out of a possible 414 points combined on Hindi and math. The number of paired observations with a swing of more than 100 points between tests (positive or negative) is about 9%. These are potentially mismatched children who were assigned the wrong ID number in the second round. However, it is impossible to determine whether there is an ID mistake or whether this is a true observation, and they are left in the sample. The results are not affected if we drop these children from the estimation sample.

4 Results

We begin our analysis by looking at the summary statistics. There is a pretty striking relationship between HAZ and test scores that can be seen in the summary statistics. Table 4 shows mean test scores for categories of anthropometric z-scores—recall that a one unit change in a z-score is equivalent to one standard deviation of the reference distribution. Larger height-for-age (HAZ) is clearly associated with higher test scores. In fact, there is an obvious upward trend in scores as HAZ increases. Given our discussion about short-term and long-term measures of nutrition and their effects on schooling performance in Section 3.1, it makes sense that BMI-for-age is not so obviously related to test scores, while height-for-age appears to have a strong statistical link to testing outcomes.

Table 4: Mean Test Scores and Nutrition, by Subject (Pooled Sample)

	Mean Hindi Score	**Mean Math Score**
All	54.35	57.51
$HAZ < -3$	30.57	30.96
$-3 < HAZ < -2$	46.68	47.48
$-2 < HAZ < -1$	54.69	58.12
$-1 < HAZ < 0$	63.96	67.45
$0 < HAZ < 1$	62.66	68.40
$1 < HAZ < 2$	64.73	72.62
$HAZ > 2$	66.30	76.18
$BAZ < -3$	54.55	56.82
$-3 < BAZ < -2$	53.11	55.81
$-2 < BAZ < -1$	56.29	59.04
$-1 < BAZ < 0$	53.29	58.11
$0 < BAZ < 1$	52.68	55.06
$1 < BAZ < 2$	56.27	60.92
$BAZ > 2$	50.50	52.91

Source: Author's calculations. Standard deviations in parentheses. Total score is the sum of math and Hindi sections.

4.1 Pooled data

We begin by showing results in Table 5 for the pooled sample that estimates the conditional demand function of Equation 12. The regressions are run for the two subjects, Hindi and mathematics, separately. Apart from the relevant teacher match, the righthand side variables for the Hindi and math regressions are identical. Because the dependent variable is normalized with mean zero and standard deviation one, the coefficients in Table 5 can be interpreted as a 1-unit change in a variable x resulting in a change of β standard deviations from the mean test score.

We use a very long list of righthand side (RHS) variables (see Table 13 in Appendix B), based on our conditional demand function.[16] The regressions include several time-varying and time-invariant explanatory variables that are not reported in the tables in the interest of brevity. These include individual characteristics, household and parental characteristics, school characteristics, classroom characteristics[17], teacher characteristics, health environment characteristics, cost of schooling and health inputs, and district dummies.[18] "Wave 2" is a time dummy taking the value of 1 for every observation in the second round of testing. It controls for the average increase in test scores over the academic year that is common to all of the students. The individual-level time-variant health variable of interest is HAZ. We also include two shock measures (S_2), days of illness in the last 3 months (a short-term measure that varies from 0 to 90) and a dummy variable for whether the child was absent three or more days from school in the last two weeks. These two variables may help to control for the child's testing ability on the given day when the enumerators visited the school.

Columns [1] and [4] show simple OLS results for Hindi and math, respectively. These columns are important to show because despite the barrage of controls included, the coefficient on HAZ is positive and significant at the 1% level. It is clear the girls score much lower than boys on average, and that the extent to which they trail boys is larger for math than Hindi. Class 4 children score higher than class 2 children, as we would expect (we refer the reader to Appendix

[16]Note that we do not control for school participation, since we specifically surveyed children who were already in school. We experimented with a Heckman selection equation, using all of the siblings observed on the household rosters, but this limited our analysis too much because we had only household level covariates for these children. Moreover, we were not confident about the identification of the selection mechanism. We note that the selection correction did not however change our results in a scaled-down, limited model. We refer the reader to Drèze and Kingdon (2001), a detailed paper on school participation in northern rural India.

[17]We do not control for class size, since unobserved school quality could increase class size. However, we note that its omission from the regressions does not alter the results in any substantial way.

[18]The survey does not contain information on parental heights. This might have been useful, in order to see whether children are shorter due to malnutrition or genetics. However, since the data are a panel, parental heights would be inconsequential in a fixed effects regression.

Table 5: Results for Pooled Sample, Selected Regressors

	Hindi			Math		
	OLS	Village FE	School FE	OLS	Village FE	School FE
	(1)	(2)	(3)	(4)	(5)	(6)
HAZ	.048	.038	.031	.078	.069	.065
	(.011)***	(.007)***	(.010)***	(.010)***	(.009)***	(.010)***
Class 4	.747	.827	.843	.681	.732	.729
	(.060)***	(.027)***	(.054)***	(.047)***	(.045)***	(.046)***
Girl	-.144	-.143	-.143	-.311	-.314	-.309
	(.029)***	(.018)***	(.028)***	(.027)***	(.027)***	(.026)***
Age (years)	.055	.049	.035	.095	.085	.079
	(.014)***	(.008)***	(.013)***	(.012)***	(.011)***	(.012)***
Days of illness (last 3 mos)	-.004	-.004	-.003	-.004	-.004	-.003
	(.0009)***	(.0009)***	(.0009)***	(.0009)***	(.0009)***	(.0009)***
Absent 3+ days (last 2 wks)	-.163	-.151	-.150	-.143	-.133	-.133
	(.028)***	(.023)***	(.025)***	(.027)***	(.025)***	(.025)***
Wave 2	.226	.238	.240	.287	.277	.286
	(.024)***	(.019)***	(.020)***	(.023)***	(.023)***	(.020)***
Mother's education (years)	.014	.013	.012	.014	.014	.013
	(.005)***	(.004)***	(.005)**	(.005)***	(.005)***	(.005)**
Father's education (years)	.017	.017	.016	.015	.014	.013
	(.003)***	(.002)***	(.003)***	(.003)***	(.003)***	(.003)***
Private school	.348	.554		.379	.422	
	(.154)**	(.094)***		(.127)***	(.149)***	
Obs.	7842	7842	7842	7842	7842	7842
Groups		133	159		133	159
R^2	.413	.372	.325	.487	.412	.372

The dependent variables are normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

C for more regression results separated by gender and class). Older children also perform better. Being absent for three days or more in the last two weeks is associated with a lower score, as is days of illness in the last three months. In addition, parents' educations of both the mother and father are very significant.

The coefficients for HAZ across the two subjects are quite similar. One more unit of HAZ increases the Hindi test score by about .05 of a standard deviation (about 2.8 achievement points—about a 5% increase at the mean score of 54.4) and the math score by about .08 of a standard deviation (about 4.5 achievement points—about an 8% increase at the mean score of 57.7).[19] Compared to Glewwe et al. (2001), who found in their 2SLS results that one standard deviation of HAZ resulted in a rise of achievement scores by one-third of a standard deviation, these results are small. (However, our IV results match up more closely to the results found in Glewwe et al. (2001)).

The results in Table 5 also suggest that the effect of HAZ may be somewhat lower for Hindi than for math; this is also confirmed if we pool the subjects (thereby having four observations per child—two for each subject) and estimate a similar school fixed effects model. However, while these coefficients are quite precisely estimated, their 95% confidence intervals do overlap.

In addition, students in private schools experience much better achievement on average than those in government schools. This is due in part to the higher quality of private schools; however, there is almost certainly some selection effect going on where students from more privileged backgrounds with parents who value education highly attend private school. Within the same village, the private school advantage is greater than across villages. That the marginal effect on the private school dummy is about double in the village fixed effects column compared with the OLS column is consistent with the idea that, in India, private schools are more likely to spring up in poorer areas and in villages where the government schools function poorly (as measured by teacher absence rates or by poor school infrastructure), as found in Muralidharan and Kremer (2006) and Pal and Kingdon (2009).

[19] The standard deviation of the Hindi (math) marks is 58 (49).

The data allow for several levels at which we can group students. We present village and school fixed effects for the pooled sample, and note that the data strongly reject random effects specifications in favour of fixed effects, as indicated by simple Hausman tests, and therefore we do not show any random effects results. The fixed effects (within) estimation results are consistent in showing a positive relationship between HAZ and test scores. The introduction of fixed effects at the village and school levels lower the coefficient on HAZ slightly, suggesting that some school-level and village-level variables are correlated with HAZ. However, HAZ remains significant at the 1% level for both Hindi and math, even with school fixed effects.

4.2 Siblings and household fixed effects

We have established that a strong statistical relationship between achievement and HAZ is present in the data. However, we must still address the significant problem of endogeneity. The primary method with which we tackle endogeneity is to narrow down the sample to siblings in the same household, who also attend the same school (most do attend the same school). Note that this creates a sub-sample in which each household has four observations—two for each sibling. This allows us to use a household fixed effects estimator that can control for time-invariant unobservables at the levels of the village, school, and most importantly, the household. This specification allows us to control for variation in parental attitudes and practices with respect to health *and* education.

Table 6 shows the results of regressions run on the siblings sub-sample for both Hindi and math. We present OLS, village fixed effects and school fixed effects for completeness and for comparison with the full, pooled sample. The first thing to note is that though the point estimates in the sub-sample regressions are similar to those in Table 5 (and are not statistically different), they are slightly higher. This may be due to the imprecision caused by a smaller sample size, or it may be because there is something different about the children that have siblings in the same school, two classes apart, as compared with the other students. This is impossible to determine, but remains a potential consideration when interpreting the results.

Focusing now on Columns [4] and [8] of Table 6, it is clear that even with the household fixed effects, there is a strong, positive effect of HAZ for both math and Hindi. In fact, going from the school level to the household level has very little effect on the estimates, and if anything pushes them up. This suggests that there are some variables, most likely parental preferences for education and health, at the level of the household that are positively correlated with both academic achievement and HAZ.

To get a further sense of the size of the effect of HAZ, the interesting calculation is the following: being in Class 4 increases the test score by roughly .9 standard deviations (SDs) in both subjects, *ceteris paribus*, so one year of school is related to an increase of about .45 SDs and one day of school is related to about .0025 SDs (= .45/180, assuming a 180-day school year, though many schools in rural India have fewer days of school). If we think one unit of HAZ (or about 6cm) raises the test score by .08 SD, then one unit if HAZ is worth about one month (32 days) of school, and an additional centimeter of height is equivalent to about 5 days of school attendance (= 32/6). Moreover, the 1.26 units of HAZ that represents the average deficit of the Indian children relative to the healthy reference population translates into a difference of over 40 days in school.[20] Since the children have spent, on average, three years in school, this is like saying that Indian children spend about 7% less time in school per year, just due to a health disadvantage. If the distribution of HAZ remains constant over the 6 years that these children spend in primary school, or in other words their growth does not "catch up" to that of American children, this can help explain why the performance of these rural children is low relative to healthier peers in India and in other countries.[21]

In addition, we can compare the "academic value" of one SD of HAZ compared to father's education. The coefficients on father's education is roughly .01 in all the regressions. The SD of father's education is 5 years. Therefore, a one SD increase in father's education is associated

[20]Mean HAZ is -1.26.

[21]While India has not taken part in international tests of learning achievement levels in the past three decades, a recent application of the TIMSS test items to secondary schools student in two north Indian states (Rajasthan and Orissa) showed that only about 15% of Indian children reached or surpassed the international TIMSS average test score; the achievement of 85% of the students was below the international mean mark (Wu, 2009).

Table 6: Results for Same-School Siblings Only, Selected Regressors

	Hindi				Math			
	OLS	Village FE	School FE	HH FE	OLS	Village FE	School FE	HH FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HAZ	.073	.062	.055	.080	.092	.073	.075	.088
	(.021)***	(.024)***	(.025)**	(.029)***	(.024)***	(.027)***	(.029)***	(.027)***
Class 4	.869	.967	.985	.883	.812	.911	.898	.888
	(.093)***	(.085)***	(.091)***	(.119)***	(.090)***	(.096)***	(.100)***	(.085)***
Girl	-.087	-.103	-.099	-.119	-.303	-.315	-.278	-.310
	(.050)*	(.052)**	(.052)*	(.057)**	(.049)***	(.050)***	(.053)***	(.053)***
Age (years)	.026	.032	.032	.060	.054	.041	.056	.074
	(.029)	(.034)	(.033)	(.034)*	(.028)*	(.030)	(.031)*	(.031)**
Days of illness (last 3 mos)	-.004	-.004	-.003	-.0003	-.006	-.005	-.004	-.0003
	(.003)*	(.003)	(.002)	(.002)	(.002)***	(.002)**	(.002)**	(.001)
Absent 3+ days (last 2 wks)	-.214	-.223	-.241	-.186	-.164	-.156	-.162	-.109
	(.056)***	(.055)***	(.054)***	(.043)***	(.056)***	(.057)***	(.055)***	(.041)***
Wave 2	.169	.182	.207	.176	.245	.250	.264	.238
	(.033)***	(.037)***	(.034)***	(.037)***	(.037)***	(.037)***	(.036)***	(.036)***
Mother's education (years)	-.005	-.004	-.003		.005	.013	.009	
	(.015)	(.019)	(.018)		(.015)	(.018)	(.019)	
Father's education (years)	.018	.015	.013		.017	.013	.009	
	(.006)***	(.007)**	(.008)*		(.007)**	(.008)*	(.008)	
Private school	.568	.516			.477	.375		
	(.207)***	(.246)**			(.155)***	(.247)		
Obs.	1436	1436	1436	1436	1436	1436	1436	1436
Groups		112	129	359		112	129	359
R^2	.491	.434	.397	.414	.525	.468	.448	.524

The dependent variables are normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

with a 5% increase in the child's Hindi and math scores. Both of these are lower than the effect of one SD of HAZ: a one unit increase in HAZ is roughly equivalent, *ceteris paribus*, to a child's father experiencing an education boost of more than 5 years.

5 Robustness checks

Though the results in Section 4 are suggestive, they can be further strengthened through several robustness checks discussed below.

5.1 Instrumental variables

To add robustness to the household fixed effects estimates, we now instrument for HAZ using rainfall data for the months of each child's birth year, age one, and age two.[22] While these instruments do not vary within child, they do vary within household across siblings who are of different ages (a similar argument is made by Alderman et al. (2006)).

We present IV results in Table 7 for both Hindi and math. Our first stage results are included in Appendix D. Overall, the IV and IVFE coefficients are higher than the OLS and FE coefficients. However, they are close enough that a test of endogeneity indicates that HAZ is exogenous.[23] The IV approach suggests that at least for Hindi, there is a significant effect of HAZ on academic performance, even after estimating household fixed effects. This effect is smaller for math, and statistically insignificant when controlling for household fixed effects.

Based on a comparison of the FE and IVFE coefficients, the OLS and FE estimates appear to be biased downwards for both subjects. All in all, the relative proximity of the estimates suggests that the endogeneity of HAZ is not causing a large bias. However, we must qualify this interpretation with an acknowledgement of weak instruments that may not offer a significant improvement over the OLS/FE results in Table 6. Although the overidentification test statistic

[22] Mis-measuring age will affect the accuracy of these instruments.

[23] The endogeneity test is a difference of two Sargan-Hansen statistics (similar to the Wu-Hausman test). The first is derived when the regressor (HAZ) is treated as exogenous, and the second is derived when the regressor is treated as endogenous.

Table 7: IV Results for Same-School Siblings Only, Selected Regressors

	Hindi				Math			
	2SLS	IV-Village FE	IV-School FE	IV-HH FE	2SLS	IV-Village FE	IV-School FE	IV-HH FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HAZ	.208	.307	.282	.250	.158	.239	.219	.129
	(.078)***	(.092)***	(.084)***	(.098)**	(.076)**	(.082)***	(.078)***	(.089)
Class 4	.784	.744	.770	.674	.769	.776	.774	.842
	(.095)***	(.115)***	(.117)***	(.163)***	(.091)***	(.115)***	(.113)***	(.137)***
Girl	-.087	-.098	-.101	-.137	-.303	-.305	-.276	-.314
	(.053)*	(.060)	(.061)*	(.062)**	(.049)***	(.054)***	(.058)***	(.053)***
Age (years)	.063	.113	.108	.126	.072	.096	.105	.090
	(.035)*	(.045)**	(.045)**	(.054)**	(.032)**	(.039)**	(.039)***	(.049)*
Days of illness (last 3 mos)	-.004	-.003	-.002	-.0004	-.006	-.004	-.004	-.0003
	(.003)	(.003)	(.003)	(.002)	(.002)***	(.002)**	(.002)*	(.001)
Absent 3+ days (last 2 wks)	-.235	-.271	-.273	-.200	-.174	-.181	-.178	-.112
	(.055)***	(.059)***	(.055)***	(.043)***	(.055)***	(.058)***	(.054)***	(.040)***
Wave 2	.065	-.014	.031	.045	.193	.122	.153	.206
	(.072)	(.089)	(.072)	(.082)	(.071)***	(.073)*	(.066)**	(.081)**
Mother's education (years)	-.003	.006	.007		.006	.020	.016	
	(.016)	(.022)	(.021)		(.016)	(.019)	(.020)	
Father's education (years)	.022	.019	.014		.019	.015	.010	
	(.007)***	(.009)**	(.009)*		(.008)**	(.008)*	(.008)	
Private school	.613	.655			.483	.404		
	(.192)***	(.243)***			(.150)***	(.242)*		
Obs.	1436	1436	1436	1436	1436	1436	1436	1436
Groups		112	129	359		112	129	359
R^2	.463	.336	.311	.377	.518	.422	.414	.522
Sargan-Hansen J-statistic	38.894	32.768	33.729	36.632	40.912	39.233	40.456	37.613
P-value	.299	.576	.529	.393	.227	.286	.242	.35
First-stage F-statistic	2.845	4.434	6.837	5.596	3.295	4.013	4.971	6.249
Shea's Partial R^2	0.0662	0.0829	0.0949	0.1130	0.0658	0.0793	0.0894	0.1111

The dependent variable is normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

suggests that the instruments are valid, the power of this test is weak with weak instruments. The tests are only valid if there are at least as many suitable instruments as there are variables that are instrumented, so it may have lower power in our case. Given the size of the first stage F-statistic, which is under the rule-of-thumb threshold of 10, our instruments are weak and could actually cause further bias (Bound, Jaeger, & Baker, 1995). Indeed, the bias of our IV results could be around 20% of the OLS estimates (Stock & Yogo, 2005).

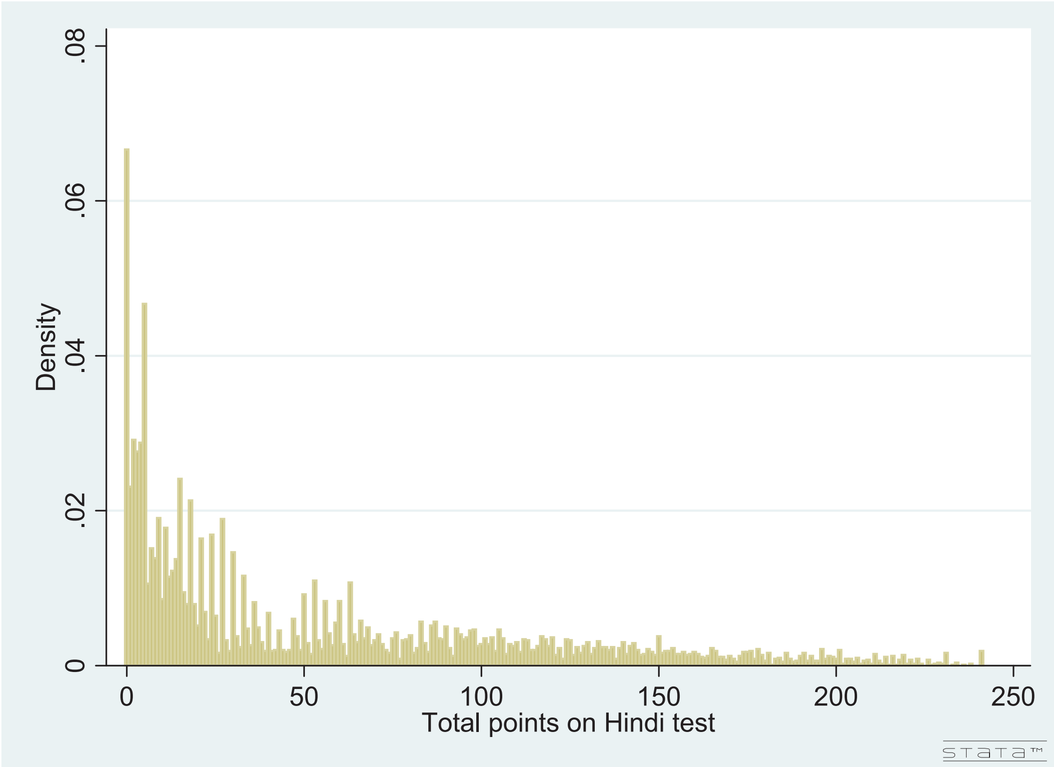
5.2 Nonlinear models

An additional criticism of the approach thus far is that the distributional qualities of the dependent variable, the test scores, may cause some efficiency problems in our estimation. Indeed, the distributions of Hindi and math scores of all the observations in both waves are shown in Figures 2 and 3. These graphs highlight two important features of the data. First, for each distribution, there is clearly a mass point at zero. Second, the distributions have a very long right tail extending to the maximum score. The distributions' shapes are not significantly altered when logged.

The interesting point to take away from these figures is that there are many children who know virtually nothing, even after several years of schooling. There are some who know a little bit, though they are still below the expected standard for their grade. A few students manage to learn at the expected level. To be sure, some of the distribution's shape comes from combining private and government (public) schools, boys and girls, Bihar and UP, and class 2 and 4, because there is a lot of heterogeneity in student achievement among these groups. For example, if we look only at class 4 students in Bihari private schools, the achievement scores have a more normal appearance, though still with relatively thick tails. However, the distributions remain very similar for students in government schools, no matter how we break up the data.

Before presenting results from the count regressions, it is important to acknowledge the trade-offs between a linear approach and a nonlinear approach. First, the linear model is often

Figure 2: Hindi Test Scores Histogram, Pooled Waves – All Children in Classes 2 and 4

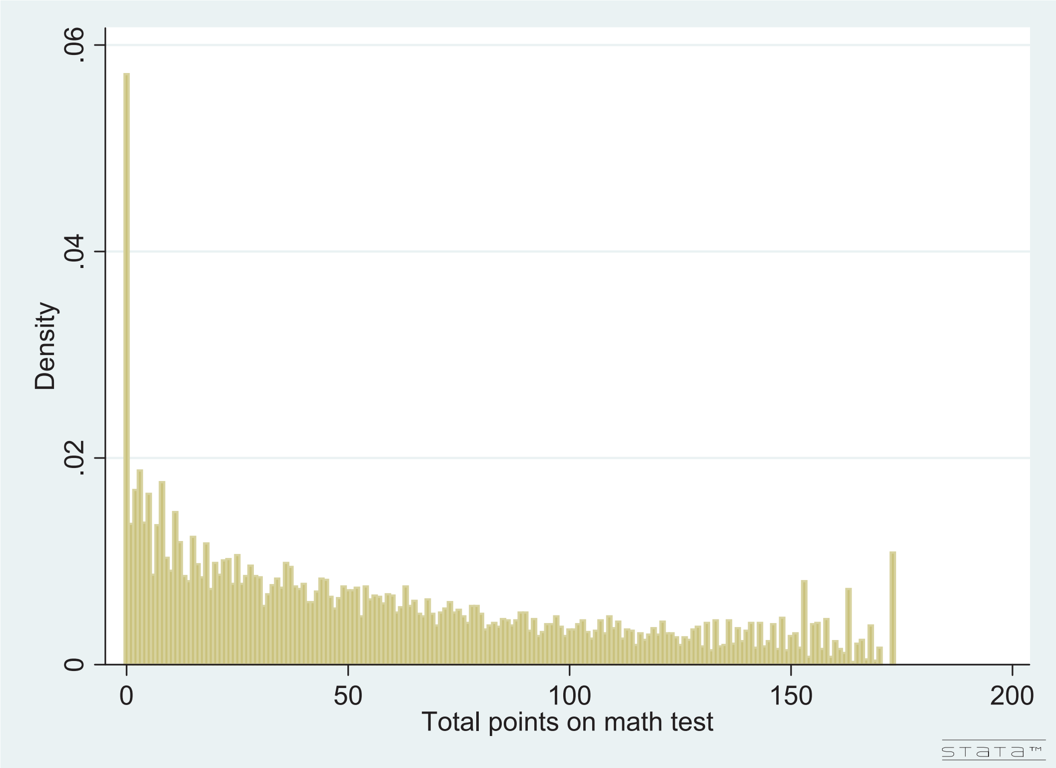


inappropriate for left-censored data, i.e. the predicted values may be negative. In addition, while the linear model allows us to investigate issues of endogeneity and fixed effects simultaneously in a familiar framework, the inherent heteroskedasticity of the distribution seen in Figure 2 means that inference may be inefficient. With 7,800 observations in the pooled sample, this may not be an issue; however, in the siblings sub-sample, there may be some gain to considering a different econometric strategy.[24]

The alternative is a nonlinear model, and in particular a count model, which can allow for the fact that tests scores are non-negative whole numbers, and may be better at tackling the problem of heteroskedasticity inherent in the test scores distribution. In contrast to the linear

[24]The standard errors are robust and clustered at the village level in all of the results tables, but homoskedasticity is strongly rejected in all of the models. This heteroskedasticity remains even with several interaction terms that try to capture the heterogeneity and nonlinearities. For example when we include interaction terms for every explanatory variable interacted with the private school dummy, there is still evidence of heteroskedasticity. When we separate the data into smaller groups by gender, state, and class, this reduces the degree of heteroskedasticity by allowing flexibility in the estimation. On the other hand, we run into small sample problems, and in addition, regressions based on the linear model for class 2 in both Bihar and UP *still* exhibit heteroskedasticity problems. We suspect that this is due to the mass point of zeros that are especially prevalent for the younger children.

Figure 3: Mathematics Test Scores Histogram, Pooled Waves – All Children in Classes 2 and 4



model $\mathrm{E}(y|\mathbf{x}) = \mathbf{x}\beta$, a count model takes the form $\mathrm{E}(y|\mathbf{x}) = \exp(\mathbf{x}\beta)$. The most basic count regression model assumes that y given $\mathbf{x}$ follows a Poisson distribution, and the conditional density of y is

$$f(y|\mathbf{x}) = \frac{\exp\left[-\mu(\mathbf{x})\right][\mu(\mathbf{x})]^y}{y!}, \qquad y = 0, 1, 2 \ldots$$

where $\mu(\mathbf{x}) \equiv \mathrm{E}(y|\mathbf{x})$ (Wooldridge, 2002).

However, the count model is not without its limitations. First, the Poisson distributional assumption that the conditional mean and variance are equal can be quite restrictive:

$$\mathrm{Var}(y|\mathbf{x}) = \mathrm{E}(y|\mathbf{x}). \tag{14}$$

This "equidispersion" assumption can cause the econometrician to underpredict the prevalence of zeros in the data. Another potential problem with the equidispersion assumption is that most count data has a variance that exceeds the mean, termed overdispersion. Our data on test scores are a perfect example of overdispersion: the means of both the math and Hindi scores are much less than their variances (they are in fact roughly the same as their standard deviations). There are several ways to resolve this issue. We can correct the standard errors by using robust inference or bootstrapping, or we can relax the assumption in Equation 14 so that

$$\mathrm{Var}(y|\mathbf{x}) = \sigma^2 \mathrm{E}(y|\mathbf{x}). \tag{15}$$

When $\sigma^2 > 1$, the variance is greater than the mean. A generalized linear model estimated by quasi-maximum likelihood (though still with the Poisson distributional assumption) allows Equation 15 to hold. Another case in which Equation 15 holds is when we use a negative binomial as the underlying distribution as opposed to Poisson. This is often referred to as NegBin I (Cameron & Trivedi, 1986). Thus we have the tools with which to handle the potential for erroneous inference caused by overdispersion. The second difficulty posed by the count model is that we are restricted to the Poisson distributional assumption when expanding the analysis to fixed effects or instrumental variables[25] because a FE or IV estimator using the negative binomial distribution does not exist. This is not a huge problem, since we know how to correct the standard errors of the Poisson estimator. Finally, with a count model, we are unable to combine FE and IV as we can with the linear model.

The results of various specifications with the count model for Hindi are presented in Table 8, and Table 9 for math. For these regressions, we use the siblings sub-sample, in order to compare them to the main estimates above. In general, the results in the count regressions match those in Tables 6 and 7 quite well for all the covariates presented. The regressions' R^2 are very close to the R^2 of the linear estimators. Note that the coefficients in the nonlinear regressions are interpreted as if the LHS variable were logged. So, in Column [1] of Table 8, the coefficient of .098 on HAZ means that the expected count score (total score) changes by a factor of 1.10 (= exp(.098)), or increases by about 10%, *ceteris paribus*. Note that this is very close to the result in Column [1] of Table 6. In the negative binomial model the marginal effect of an additional unit of HAZ when evaluated at the mean is about 7 (8) points for Hindi (math), and it differs by class: the marginal effect of HAZ at the mean is 4.5 (6) points for children in Class

[25]See Mullahy (1997) for a discussion of count data and IV.

Table 8: Count Regressions for Hindi (Siblings Only): Poisson IV and Fixed Effects (FE), Selected Regressors

	Poisson	NegBin I	Poisson IV	Poisson Village FE	Poisson School FE	Poisson Household FE
	(1)	(2)	(3)	(4)	(5)	(6)
HAZ	.098	.153	.271	.087	.083	.135
	(.026)***	(.027)***	(.096)***	(.029)***	(.031)***	(.037)***
Girl	-.083	-.133	-.135	-.111	-.094	-.084
	(.056)	(.069)*	(.063)**	(.055)**	(.057)	(.068)
Class 4	.951	1.062	.949	1.062	1.141	.964
	(.103)***	(.108)***	(.114)***	(.099)***	(.113)***	(.150)***
Age (years)	.042	.081	.139	.051	.050	.100
	(.031)	(.035)**	(.040)***	(.035)	(.036)	(.043)**
Days of illness (last 3 mos)	-.005	-.005	-.003	-.004	-.003	-.0006
	(.004)	(.003)	(.005)	(.003)	(.003)	(.003)
Absent 3+ days (last 2 wks)	-.302	-.376	-.378	-.281	-.311	-.213
	(.067)***	(.076)***	(.085)***	(.066)***	(.063)***	(.056)***
Wave 2	.182	.313	.235	.204	.231	.186
	(.041)***	(.051)***	(.086)***	(.043)***	(.045)***	(.046)***
Mother's education (years)	-.009	-.0005	.003	-.0008	-.002	
	(.017)	(.017)	(.016)	(.021)	(.021)	
Father's education (years)	.030	.032	.034	.028	.029	
	(.007)***	(.009)***	(.008)***	(.009)***	(.009)***	
Private school	.779	.850	.918	.772		
	(.210)***	(.224)***	(.246)***	(.273)***		
Obs.	1436	1436	1436	1436	1436	1432
R^2	.49	.383	.433	.389	.155	.163
Log likelihood	-23446.88	-6635.165		-19108.16	-18575.85	-10159.74
$\hat{\alpha}$		.885				

Standard errors in parentheses are robust to heteroskedasticity. Standard errors in column [3] are bootstrapped to 50 replications. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. The number of observations falls with FE Poisson because the likelihood function drops observations that have no variation in the dependent variable across waves. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively. R^2 is calculated as the square of the correlation between $score_i$ and $\hat{score}_i$.

Table 9: Count Regressions for Math (Siblings Only): Poisson IV and Fixed Effects (FE), Selected Regressors

	Poisson	NegBin I	Poisson IV	Poisson Village FE	Poisson School FE	Poisson Household FE
	(1)	(2)	(3)	(4)	(5)	(6)
HAZ	.098	.141	.114	.084	.086	.121
	(.024)***	(.025)***	(.083)	(.027)***	(.029)***	(.028)***
Girl	-.261	-.296	-.288	-.275	-.249	-.279
	(.042)***	(.058)***	(.038)***	(.044)***	(.048)***	(.050)***
Class 4	.705	.705	.696	.815	.823	.772
	(.081)***	(.089)***	(.079)***	(.089)***	(.097)***	(.096)***
Age (years)	.063	.109	.118	.059	.068	.095
	(.025)**	(.029)***	(.032)***	(.027)**	(.029)**	(.031)***
Days of illness (last 3 mos)	-.005	-.005	-.006	-.003	-.003	.001
	(.002)***	(.002)***	(.002)**	(.002)*	(.002)*	(.002)
Absent 3+ days (last 2 wks)	-.167	-.215	-.212	-.171	-.169	-.105
	(.054)***	(.062)***	(.060)***	(.056)***	(.054)***	(.042)**
Wave 2	.202	.262	.290	.217	.230	.196
	(.034)***	(.045)***	(.072)***	(.034)***	(.035)***	(.034)***
Mother's education (years)	-.003	.013	.012	.012	.010	
	(.013)	(.017)	(.012)	(.016)	(.016)	
Father's education (years)	.018	.020	.019	.015	.012	
	(.007)**	(.008)**	(.006)***	(.007)**	(.008)	
Private school	.488	.503	.482	.492		
	(.135)***	(.166)***	(.169)***	(.207)**		
Obs.	1436	1436	1436	1436	1436	1436
R^2	.514	.433	.490	.404	.269	.226
Log likelihood	-18548.84	-6833.489		-15175.8	-14845.75	-8441.984
$\hat{\alpha}$		.657				

Standard errors in parentheses are robust to heteroskedasticity and are clustered at the village level. Standard errors in column [3] are bootstrapped to 50 replications. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. The number of observations falls with FE Poisson because the likelihood function drops observations that have no variation in the dependent variable across waves. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively. R^2 is calculated as the square of the correlation between $score_i$ and $\hat{score}_i$.

2 and 13 (12) points for children in Class 4.

The negative binomial model in Column [2] of Tables 8 and 9 finds somewhat higher effects, suggesting that there is some overdispersion in the data. The NegBin estimator may be preferable to the Poisson, as it allows explicitly for overdispersion in the distribution. A likelihood ratio test of overdispersion is given by the parameter $\hat{\alpha}$. The null hypothesis of equidispersion is $\alpha = 0$; here, the data strongly reject the null. Also, the log likelihood of the NegBin model is much higher (less negative) than the Poisson, meaning that the negative binomial distribution fits the data more precisely than the Poisson. The Poisson (NegBin I) regression can also be estimated with a generalized linear model with family Poisson (negative binomial I).[26]

We also report Poisson IV results using a GMM estimator, developed by Nichols (2007) and based on work by Mullahy (1997).[27] One criticism of the estimator is that it can generate confidence intervals that are too narrow, and may require bootstrapped standard errors. In our case, we have bootstrapped to 50 replications. Even though we favour the fixed effects approach, the IV is helpful in signing the bias on HAZ in the pooled regression. Earlier, we found evidence of a downward bias in the OLS estimator as compared with the IV.

The results of a Poisson fixed effects model are presented in Columns [4-6] of Tables 8 and 9. For the fixed effects, we use a quasi-maximum likelihood estimator that allows for robust standard errors to be calculated (Wooldridge, 2002), which is consistent even when the Poisson distribution is misspecified.[28] There is no equivalent FE model using the negative binomial distribution (Allison & Waterman, 2002).[29] We are thus using a partially parametric approach, since we are not fully specifying the variance function (Cameron & Trivedi, 1999). This lets us sidestep the typical overdispersion problems with the Poisson estimator (highlighted by the

[26]We also ran the same regression in Columns [1] and [2] of Tables 8 and 9 using a GLM estimator that corrects for overdispersion by adjusting (scaling) the standard errors for $\sigma > 1$—this does not change the coefficients. We do not report GLM standard errors because they are very similar to the robust standard errors reported in the table and do not change the interpretation of the results in any way.

[27]The Poisson IV is implemented using the `ivpois` command in Stata 10.0 written by Austin Nichols (Nichols, 2007).

[28]We use the command `xtpqml` in Stata 10.0.

[29]Using individual dummies to capture FE as an alternative is not feasible due to large number of children and the excessive degrees of freedom needed to estimate the model.

superiority of Column [2] over [1]). The Poisson FE results add further evidence to our claim that HAZ has a positive, causal impact on achievement. These results are very precisely estimated, even after correcting for overdispersion, which typically raises the nominal Poisson standard errors. Their magnitudes are also consistent with the results of the linear model, but because the estimators are better matched to the data's distribution, we obtain more precision.

The household fixed effects can be found in Column [6] for both Hindi and math. This regression uses the Poisson FE estimator at the household level and is robust to heteroskedasticity and is clustered at the village level. We find a large, statistically significant positive effect of HAZ. Using FE at the household level raises the point estimate, providing yet more evidence of a downward bias (though the size of the bias is still quite small). We surmise that the downward bias is probably mostly attenuation bias, caused mainly by measurement error.

Similar to the linear model, we find that the average test scores go up for all of the children, and that recent absences from school lower test scores. We point out once again that the size of the HAZ coefficient is large relative to the other variables, implying that nutrition is one of the biggest measurable determinants of test performance.

Although we ameliorate the overdispersion problem, we are still left with the issue of excess zeros. Zero-inflated models and hurdle models can address this problem, but lack any IV or FE applications. Therefore we present these results in Appendix E for completeness. In fact, the results from the zero-inflated Poisson and NegBin I are very similar to our results in this Section, and supply further evidence that HAZ is a key driver of test scores. Moreover, we show that HAZ is negatively associated with a score of zero. The hurdle models also agree, though they suggest a higher coefficient for HAZ and thus an even higher impact of long-term nutrition on academic achievement. We leave a deeper exploration of hurdle models for future work.

5.3 BAZ and H_2

An alternative specification to Equation 13 is to include H_2, if we think that current nutritional status could have an effect on test scores as well.[30] Though there is no reason to think that short-term fluctuations in nutrition in period 2 has an effect on cognitive development, it may still affect a child's ability to do well in school. For example, children with short-run nutritional deficiencies are probably less able to pay attention and concentrate; more likely to stay home from school; and might have less energy for lessons, learning and homework.

OLS, fixed effects, and IV results are shown in Tables 10 for Hindi and 11 for math. The evidence on BAZ is mixed. In Columns [1-4] in each table, there seems to be some evidence that both HAZ and BAZ matter, at least once household fixed effects are controlled for. For Hindi, the inclusion of BAZ reduces the coefficient on HAZ by about 0.02, but the coefficient on HAZ for math is still around 0.08. However, we cannot reject that the coefficients on HAZ are statistically the same both with and without BAZ as an additional control.

In Columns [5-8] in Tables 10 and 11, we instrument for both HAZ and BAZ. For BAZ, we use health environment characteristics: time to health providers (chemist, private doctor, health clinic, and hospital) collected in the village questionnaire, as well as two district-level tuberculosis statistics from India's Ministry of Health and Family Welfare on TB detection rate and the TB testing rate.[31] These characteristics should be valid exclusions, since we condition on H_2 and H_1. The first four instruments are constant across villages and through time, so they are only used in Columns [1] and [5]. The tuberculosis statistics are at the district level, they vary

[30] As a measure of short-term health, BAZ is preferred to weight-for-height (WHZ), as explained by both the Centers for Disease Control and the World Health Organization (Flegal, Wei, & Ogden, 2002). In fact, the use of WHZ is no longer recommended for children over the age of 5, and the available anthropometric software does not calculate z-scores for weight-for-height (see http://www.who.int/growthref/en/). This is primarily because BAZ allows an adjustment of the weight-for-height concept that takes into account age (Cole, Flegal, Nicholls, & Jackson, 2007). In contrast, WHZ is age-independent in its calculation, so comparing across children of several ages is problematic because for a given height, median weight will differ by age and sex. In addition, WHZ is expected to change seasonally and is therefore difficult to use over longer periods. However, BAZ cut-offs for "thinness" according to BAZ vary with age and sex until one reaches adulthood, so interpretation of the absolute values across ages can be difficult.

[31] The TB statistics are from the years 2006 and 2007; the testing data are from 2007 and 2008, so the match is not perfect. However, it is reasonable that the effects of public health improvements or changes would take time to have a measurable effect on public health.

Table 10: Results for Hindi Tests Scores with Same-School Siblings Only and BAZ Included, Selected Regressors

					IV			
	OLS	Village FE	School FE	HH FE	2SLS	IV Village FE	IV School FE	IV HH FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HAZ	.070	.053	.060	.065	.197	.386	.305	.339
	(.016)***	(.025)**	(.025)**	(.028)**	(.109)*	(.156)**	(.123)**	(.140)**
BAZ	.016	.002	-.008	.054	-.032	.036	-.142	-.108
	(.018)	(.030)	(.029)	(.024)**	(.163)	(.221)	(.194)	(.141)
Class 4	.855	.936	.989	.834	.776	.642	.781	.511
	(.063)***	(.088)***	(.100)***	(.109)***	(.088)***	(.149)***	(.119)***	(.181)***
Girl	-.101	-.107	-.086	-.123	-.108	-.107	-.098	-.160
	(.040)**	(.056)*	(.059)	(.064)*	(.041)***	(.049)**	(.048)**	(.052)***
Age (years)	.032	.030	.019	.061	.062	.136	.088	.154
	(.020)	(.036)	(.035)	(.031)**	(.038)	(.056)**	(.041)**	(.056)***
Days of illness (last 3 mos)	-.004	-.005	-.004	-.0005	-.003	-.003	-.003	-.0001
	(.003)	(.003)	(.003)*	(.002)	(.003)	(.003)	(.003)	(.002)
Absent 3+ days (last 2 wks)	-.216	-.240	-.262	-.168	-.233	-.306	-.301	-.200
	(.051)***	(.056)***	(.062)***	(.045)***	(.054)***	(.068)***	(.060)***	(.055)***
Wave 2	.194	.205	.218	.191	.100	-.055	.047	.001
	(.042)***	(.040)***	(.035)***	(.033)***	(.097)	(.134)	(.100)	(.113)
Private school	.478	.486			.499	.713		
	(.176)***	(.253)*			(.187)***	(.286)**		
Obs.	1424	1424	1424	1424	1424	1424	1424	1424
Groups		108	125	348		108	125	348
R^2	.481	.426	.366	.418	.455	.237	.248	.291
Sargan-Hansen J-statistic					19.909	11.361	12.949	11.98
P-value					.224	.498	.373	.447
First-stage F-statistic for HAZ					3.45	3.96	8.97	4.03
Shea Partial R^2 for HAZ					0.0828	0.0855	0.0987	0.1024
First-stage F-statistic for BAZ					2.99	4.07	6.05	4.93
Shea Partial R^2 for HAZ					0.0503	0.0583	0.0643	0.0749

The dependent variable is normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

Table 11: Results for Math Tests Scores with Same-School Siblings Only and BAZ Included, Selected Regressors

					IV			
	OLS	Village FE	School FE	HH FE	2SLS	IV Village FE	IV School FE	IV HH FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HAZ	.094	.065	.076	.082	.287	.289	.240	.033
	(.015)***	(.028)**	(.027)***	(.026)***	(.105)***	(.146)**	(.123)*	(.118)
BAZ	.038	.021	.020	.066	.052	.104	-.040	.233
	(.018)**	(.029)	(.031)	(.021)***	(.165)	(.191)	(.162)	(.110)**
Class 4	.802	.894	.907	.854	.680	.721	.791	.910
	(.059)***	(.097)***	(.095)***	(.090)***	(.092)***	(.132)***	(.107)***	(.144)***
Girl	-.297	-.305	-.255	-.305	-.308	-.290	-.258	-.288
	(.039)***	(.054)***	(.059)***	(.058)***	(.042)***	(.046)***	(.044)***	(.044)***
Age (years)	.058	.042	.048	.088	.114	.119	.096	.076
	(.020)***	(.032)	(.032)	(.030)***	(.038)***	(.054)**	(.042)**	(.048)
Days of illness (last 3 mos)	-.005	-.004	-.004	.0004	-.004	-.003	-.004	.0003
	(.002)**	(.002)**	(.002)*	(.001)	(.003)	(.002)	(.003)	(.002)
Absent 3+ days (last 2 wks)	-.177	-.178	-.184	-.107	-.206	-.208	-.203	-.093
	(.048)***	(.058)***	(.056)***	(.042)**	(.054)***	(.058)***	(.052)***	(.046)**
Wave 2	.269	.266	.279	.251	.118	.089	.161	.270
	(.040)***	(.041)***	(.037)***	(.030)***	(.092)	(.118)	(.096)*	(.095)***
Private school	.478	.580			.513	.697		
	(.148)***	(.225)**			(.177)***	(.247)***		
Obs.	1424	1424	1424	1424	1424	1424	1424	1424
Groups		108	125	348		108	125	348
R^2	.511	.461	.424	.53	.451	.362	.376	.493
P-value					.018	.627	.503	.673
Sargan-Hansen J-statistic					30.05	9.874	11.302	9.352
First-stage F-statistic for HAZ					4.19	4.06	7.04	5.56
Shea Partial R^2 for HAZ					0.0828	0.0801	0.0913	0.1025
First-stage F-statistic for BAZ					2.84	3.63	5.10	5.92
Shea Partial R^2 for HAZ					0.0504	0.0515	0.0610	0.0727

The dependent variable is normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

over time and therefore vary within a given household's set of observations. Thus the identification in the rest of the regressions comes from the TB statistics, which vary across time, but not across individuals within the same district. Therefore we expect, as with the weather data, the TB stats to be weak instruments. Nonetheless, we present the IV results in order to offer an alternative specification that is worth reporting.

The IV results for HAZ remain very consistent with those in Section 5.1. HAZ is significant in the Hindi IV-HH fixed effects regression, but not in the math IV-HH fixed effects. Once instrumented, BAZ is insignificant for Hindi test scores, but is significant for math (Column [8]). This suggests that both short-term and long-term nutrition matter for academic achievement in math and reading/writing, and that their effects may be underestimated without instrumenting.

5.4 Child-level fixed effects

A further extension of our analysis is to see how the results fare when using child-level fixed effects. This is a final examination of the robustness of our results. Advancing on the siblings analysis, fixed effects at the individual should remove any remaining characteristics of each child that do not change over the period of observation. Furthermore, if we consider the individual-level fixed effects specification as an equation in first differences, then the narrative alters slightly from one of the impact of height/nutrition on academic performance to one of physical growth and academic performance changes over time. While this question of intertemporal change is not the primary focus of this paper, it is nonetheless interesting.

Table 12 presents results for fixed effects at the child level for both linear and nonlinear specifications. As is evident in the table, the results in the linear model break down as the standard errors rise. On the other hand, the Poisson FE model shows that significant, positive coefficients on HAZ do hold up at the child level in a nonlinear model. This may be a product of the difficulty of the FE estimator to fit the data under linear constraints. Overall, these results are consistent with what we have seen in earlier sections of the paper.

Table 12: Child-level Fixed Effects, Linear and Nonlinear Models, Selected Regressors

	Hindi		Math	
	Linear Child FE	Poisson Child FE	Linear Child FE	Poisson Child FE
	(1)	(2)	(3)	(4)
HAZ	.002	.061	.013	.075
	(.023)	(.030)**	(.025)	(.028)***
Days of illness (last 3 mos)	-.0002	.0003	-.0004	-.0001
	(.0005)	(.0007)	(.0006)	(.0006)
Absent 3+ days (last 2 wks)	-.063	-.049	-.047	-.015
	(.018)***	(.019)**	(.020)**	(.019)
Wave 2	.250	.220	.315	.218
	(.027)***	(.029)***	(.026)***	(.024)***
Obs.	7842	7648	7842	7668
Groups	3921	3824	3921	3834
R^2	.167	.054	.254	.052

The dependent variables in Columns [1] and [3] are normalized to have a mean of 0 and standard deviation of 1; the dependent variables in Columns [2] and [4] are raw scores. Standard errors in parentheses are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively. R^2 is calculated as the square of the correlation between $score_i$ and $\hat{score}_i$.

6 Summary and conclusion

India has a long way to go to improve schooling outcomes for its rural population, and it is important for policymakers to understand ways in which education reform can be most effective. The hurdles faced by the Indian school system are several: inadequate schooling facilities and insufficient resources like textbooks, teacher training deficiencies, sub-standard teacher competence to deliver the curriculum (Banerji & Kingdon, 2009), high teacher absenteeism, and more.

This paper is a new step in the debate surrounding the role of health in schooling outcomes in developing countries. We use new and unique data from rural India to examine how health affects academic achievement. Due to the unusual distribution of the test score data, we present linear and nonlinear models, all of which reach the same conclusion. In addition, fixed effects at the child level and instrumental variables applications offer convincing evidence that health is an important determinant of test scores for children living in rural areas of northern India. Our unique panel data allow us to address endogeneity issues caused by omitted variables and measurement error, and we find some evidence of endogeneity and attenuation bias of height-for-age, with limited downward bias resulting. Using several specifications and an exhaustive list of explanatory variables, we present a lot of evidence that long-term health, as measured by HAZ, has a significant positive effect on schooling achievement. We also find some evidence that short-term nutrition, measured with BAZ, is important too, at least for mathematics.

There is evidence that household literacy, maternal education, orphanhood, and the quality of local health infrastructure and services are all important determinants of child health in developing countries (Gibson, 2001; Lavy, Strauss, Thomas, & de Vreyer, 1996; Thomas, Strauss, & Henriques, 1991; Thomas, Lavy, & Strauss, 1991; Sandiford, Cassel, Montenegro, & Sanchez, 1995; Beegle, De Weerdt, & Dercon, 2006). Maternal education is a promising way to improve child health (Glewwe, 1999; Rosenzweig & Wolpin, 1994), as are improvements in sanitation facilities (Lee, Rosenzweig, & Pitt, 1997), family planning (Rosenzweig & Wolpin, 1986) and health knowledge in general (Jalan & Ravallion, 2003). Moreover, insurance for rural families may be an additional way to secure against especially detrimental nutrition shocks during the early years of childhood (Alderman et al., 2001; Hoddinott & Kinsey, 2001; Dercon & Krishnan, 2000; Dercon & Hoddinott, 2003). In India specifically, Borooah (2005) finds that in rural areas, safe drinking water and village health infrastructure are all positively correlated with HAZ; in addition, Brennan, McDonald, and Shlomowitz (2004) conclude that promoting better feeding of infants for Indian mothers may also be a way forward. The evidence as a whole does suggest that there may be several policy tools with which to impact on health, and thereby schooling outcomes, for Indian children. Our results imply that efforts to improve children's nutrition may have payoffs far beyond individual and public health.

A Data appendix

The data used in this paper were collected by the Centre for the Study of African Economies (CSAE), Oxford, in 2007-2008. Rukmini Banerji of Pratham, an Indian NGO specializing in primary education in India, contributed to the design of the testing tools. The survey was funded by the Spencer Foundation. A data collection firm based in Patna hired the enumerators and was responsible for the data entry. Enumerators were trained by Geeta Kingdon of Oxford and the University of London and Rukmini Banerji of Pratham.

The School-Tells study was designed to learn about schools, teachers, and overall achievement in some of the most educationally disadvantaged parts of rural India. We first ran a pilot study in February and March 2007 in Bihar to vet the testing tools and the questionnaires. Building on the pilot, we then ran the full survey in the field from July 2007 to April 2008 (corresponding to the north Indian academic year). The survey was administered in 11 districts, five in Uttar Pradesh (UP) and six in Bihar.[32] Although we initially chose five districts in Bihar (as in UP), and thus started the survey in five districts of Bihar, one had to be abandoned due to flooding part way through the first visit, and another district which was not flood-affected was chosen as a replacement. However, between these two districts we only sampled the same number of schools as for any one of the rest of the districts. The districts were chosen to represent the different geographical regions of the two states. Within each district 3 blocks were chosen randomly and within each block four villages were sampled randomly from the census list of villages with a population of 600-5,000. In each sample village, we included one government and one private school. If total enrollment in the government school was less than 100, we chose a nearby village as a replacement. If there was no private school in the sample village, we replaced it with a government school in a nearby village. We surveyed 16 schools per district, making a total of 80 sample schools per state, and 160 schools overall.

The study consisted of several questionnaires designed to measure all aspects of the children's

[32] Districts are: Agra, Bijnor, Lucknow, Mahoba, Shrawasti, Banka, Bhojpur, Gaya, Madhepura, Madhubani, and Saran.

learning environment. Altogether, we collected information at the child, household, teacher, school, and village levels. The enumerators visited the schools four times, at roughly equal intervals, over the nine months of the school year 2007-08, and collected school level information during each of these visits. The testing of the children occurred in the first visit and the last visit. Anthropometric data and short-term health measures were collected on the same day as the test, at the start and at the end of the school year (nine months apart). Detailed household, village and teacher questionnaires were also administered, but most of the information on these was collected only on the first visit to each school. Within the school, enumerators visited class 2 and class 4, took attendance, and then chose children randomly from the list of attendees.

Enumerators administered the tests directly to the children. Some of the testing, such as the recognition of letters and numbers, required face-to-face interaction. Enumerators were trained to have a standardized scoring system, which applied to all children equally. At the schools, headmasters or acting principals were interviewed, and the enumerators collected information about the school infrastructure based on personal observation, e.g. the presence of a boundary wall, etc. Household heads were interviewed for the household questionnaire. This questionnaire was deliberately shortened from the pilot to the full survey.

The total number of observations in the data is 11,253. After the HAZ variable was cleaned (see Section A.1 below), the number of observations with matched health and test score information is 9,008. After dropping observations that do not have a second period match and a teacher match, we are left with with 7,842 paired observations with test scores and HAZ observed in both periods.

A.1 Height and weight

We calculate anthropometric z-scores using the 2000 CDC growth charts (O'Donnell et al., 2007).[33] These are implemented using `zanthro` in Stata 10.0. There is some evidence that using the revised 2006 WHO growth charts is preferred for developing countries because of a

[33] Ogden et al. (2002) offers a clear explanation of the advantage of the 2000 version over the original 1977 growth statistics.

better underlying sample from which the statistics are derived. However, there are two reasons why we use the CDC charts. First, the WHO standards are most relevant for early infancy. Indeed, the WHO standards were revised to target children under five (for example the WHO Anthro 2005 software only goes up to 1856 days of age), while the children in our sample are ages 6 and above. Second, the CDC and WHO standards for HAZ are actually very similar for children over 5, implying that using the CDC reference population as opposed to the WHO should not make a substantial difference for the conclusions reached in this paper. In fact, if anything, we underestimate the degree of stunting by using the CDC charts, as the American children are on average shorter than the children in the WHO sample (de Onis, Garza, Onyango, & Borghi, 2007).

The data required some cleaning of the anthropometric measurements. In particular, we set height to missing for children who were shorter (in terms of raw height) in period 2 than in period 1, or who grew more than 15 centimeters—clearly such cases are caused by either measurement error, since the amount of time between periods was only nine months, or by ID mismatches. We also set weight to missing if the children gained or lost more than 10kg. Given our sample restrictions discussed in Section 3.4, this effectively means that these children are dropped from the sample used in this paper.

These measurement errors were less likely to occur for taller children, children who score higher on the tests, and those who live in the state of Bihar. This is not likely to be a function of enumerator training, as they were all trained in the same way. Having a height measurement missing in period 2 is not a function of one's height in period 1 after controlling for gender, age, and state. Addressing the selection bias caused, if any, by such mis-measurement is beyond the scope of this paper, as identification would be extremely difficult. It may cause us to underestimate effect of height on test scores, as we are leaving out children with disproportionately low scores—those who are younger and who live in UP. Altogether, this cleaning forces us to drop about 8% of the observations in period 2. Following suggestions by the WHO, we also put

HAZ (BAZ) to missing if it is below -6 (-5) or above 6 (5).[34]

[34] See http://www.who.int/nutgrowthdb/software/Differences_NCHS_WHO.pdf. The World Bank recommends a narrower range for HAZ (-5 to 3); we do not adopt this narrower range, but even if we did, the restriction would eliminate only about 1% of the sample. Therefore it has no impact on our analysis.

B Extra tables

Table 13: Summary Statistics, Righthand Side Variables

Variable	Mean	Std. Dev.	Min.	Max.	N
Individual, time-invariant					
Girl	0.464	0.499	0	1	7924
Class 4	0.505	0.5	0	1	7924
Age	9.036	1.528	7	14	7924
Oldest	0.307	0.461	0	1	7924
Only child	0.134	0.341	0	1	7924
Individual, time-variant					
HAZ	-1.273	1.371	-5.999	4.618	7924
BAZ	-1.57	1.054	-4.995	1.633	7625
Days of illness (last 3 mos)	4.823	9.829	0	90	7924
Absent 3+ days	0.195	0.396	0	1	7924
Wave 2	0.5	0.5	0	1	7924
Household, time-invariant					
Banka	0.075	0.263	0	1	7924
Bhojpur	0.047	0.212	0	1	7924
Bijnor	0.128	0.334	0	1	7924
Gaya	0.083	0.276	0	1	7924
Lucknow	0.103	0.304	0	1	7924
Madhepura	0.118	0.322	0	1	7924
Madhubani	0.026	0.158	0	1	7924
Mahoba	0.12	0.325	0	1	7924
Saran	0.108	0.31	0	1	7924
Shrawasti	0.082	0.274	0	1	7924
Reference category: Agra					
Mother reads	0.131	0.337	0	1	7924
Father reads	0.218	0.413	0	1	7924
Note: These reading statistics are low because missing values are replaced with zero.					
Mother's education	1.613	3.211	0	18	7924
Father's education	5.020	4.952	0	18	7924
Note: These education levels are low because missing values are replaced with zero.					
Mother's education missing	0.053	0.225	0	1	7924
Father's education missing	0.063	0.243	0	1	7924
Muslim	0.104	0.306	0	1	7924
ST/SC caste	0.282	0.45	0	1	7924
OBC caste	0.596	0.491	0	1	7924
BPL card	0.495	0.5	0	1	7924
Non-productive assets	5.583	5.51	0	35.5	7924
Electricity	0.14	0.347	0	1	7924
Wage employment	0.048	0.214	0	1	7924
Manufacturing	0.117	0.321	0	1	7924
Farmer	0.271	0.445	0	1	7924
Productive assets	1.943	1.792	0	11	7924
School, time-invariant					
Private	0.118	0.322	0	1	7924
School, time-variant					
School time table	0.618	0.486	0	1	7924
Number of rooms	2.514	1.304	0	7	7924
Verandah	0.479	0.5	0	1	7924
Library	0.59	0.492	0	1	7924
Class, time-variant					
Teacher table	0.915	0.279	0	1	7924
Fan	0.021	0.142	0	1	7924
Window opens	0.647	0.478	0	1	7924
Blackboard	0.859	0.348	0	1	7924
Hindi teacher, time-invariant					
10 years of education and below	0.078	0.268	0	1	7842
12 years of education	0.413	0.492	0	1	7842
BA	0.357	0.479	0	1	7842
Reference category: MA					

BTC	0.247	0.431	0	1	7842
BEd	0.037	0.188	0	1	7842
Reference category: Any other training or none					
Regular teacher	0.303	0.46	0	1	7842
Parateacher	0.621	0.485	0	1	7842
Reference category: Headmasters & acting headmasters					
Male	0.525	0.499	0	1	7842
Experience	6.567	8.323	0	56	7842
Age	32.584	10.669	17	62	7842
Maths teacher, time-invariant					
10 years of education and below	0.06	0.237	0	1	7842
12 years of education	0.347	0.476	0	1	7842
BA	0.425	0.494	0	1	7842
Reference category: MA					
BTC	0.219	0.413	0	1	7842
BEd	0.039	0.194	0	1	7842
Reference category: Any other training or none					
Regular teacher	0.34	0.474	0	1	7842
Parateacher	0.601	0.49	0	1	7842
Reference category: Headmasters & acting headmasters					
Male	0.648	0.478	0	1	7842
Experience	6.545	7.417	0	39	7842
Age	32.106	10.046	17	62	7842
Price of schooling, time-invariant					
Distance to school (minutes)	10.195	7.192	0	60	7924
Free books	0.629	0.483	0	1	7924
Free uniform	0.027	0.161	0	1	7924
Scholarship	0.024	0.155	0	1	7924
HH spending on education per child	860.498	693.675	0	7540	7924
School fees	79.605	186.308	0	1800	7924
Instruments for HAZ					
jan-rain-birthyear	15.907	10.084	0	65.19	7924
feb-rain-birthyear	12.795	10.936	0	44.34	7924
mar-rain-birthyear	6.588	7.34	0	41.62	7924
apr-rain-birthyear	6.355	6.009	0	29.45	7924
may-rain-birthyear	35.705	34.015	0.34	147.19	7924
jun-rain-birthyear	89.958	44.546	14.01	279.77	7924
jul-rain-birthyear	210.565	81.097	42.59	594.310	7924
aug-rain-birthyear	260.582	77.863	94.180	476.53	7924
sep-rain-birthyear	166.744	57.482	29.61	382.84	7924
oct-rain-birthyear	42.867	47.54	0.2	256.56	7924
nov-rain-birthyear	11.6	17.4	0	84.63	7924
dec-rain-birthyear	5.29	8.512	0	37.98	7924
jan-rain-age1	11.607	9.525	0	65.19	7924
feb-rain-age1	13.19	10.782	0	44.34	7924
mar-rain-age1	8.088	8.418	0	41.62	7924
apr-rain-age1	8.817	6.972	0	28.85	7924
may-rain-age1	33.811	31.387	0.34	147.19	7924
jun-rain-age1	91.118	49.114	14.01	279.77	7924
jul-rain-age1	211.492	71.181	42.59	440.37	7924
aug-rain-age1	246.128	79.711	94.180	476.53	7924
sep-rain-age1	150.493	68.180	29.61	311.43	7924
oct-rain-age1	55.626	63.758	0.2	256.56	7924
nov-rain-age1	6.953	12.207	0	84.63	7924
dec-rain-age1	1.907	5.139	0	37.98	7924
jan-rain-age2	14.228	9.043	0	44.28	7924
feb-rain-age2	12.554	10.641	0	44.34	7924
mar-rain-age2	7.567	9.134	0	41.62	7924
apr-rain-age2	6.955	6.926	0	22.05	7924
may-rain-age2	28.553	26.975	0.79	147.19	7924
jun-rain-age2	83.850	44.099	14.01	226.03	7924
jul-rain-age2	204.352	83.345	26.96	440.37	7924
aug-rain-age2	235.968	68.802	94.180	476.53	7924
sep-rain-age2	150.266	69.346	39.12	304.29	7924
oct-rain-age2	48.243	43.632	0.2	256.56	7924
nov-rain-age2	5.376	10.331	0	84.63	7924
dec-rain-age2	2.34	4.47	0	37.98	7924

Instruments for BAZ					
Time to chemist	28.503	30.894	0	150	7924
Time to doctor	46.978	37.389	1	180	7924
Time to clinic	33.446	32.505	0	150	7924
Time to hospital	58.492	44.414	0	240	7924
TB detection rate	43.506	16.587	12	73	7924
TB treatment rate	414.21	214.587	169.833	915.195	7924

C Gender and class differences

In this section, we investigate the differences between genders and classes, as it is an interesting question whether the effects of HAZ on Hindi and math vary along these two important dimensions. In terms of gender, girls are on average more stunted than boys—a t-test of the difference in means across genders shows there is a significant difference (p-value= 0.0009)—which is consistent with the general evidence on gender equity in developing countries. It is possible that the effect of HAZ could depend on gender, if cognitive development differs between boys and girls, as it reasonably could. In addition, since the classes are comprised of children at different ages who may be learning a different set of skills, long-term health could be more or less important for the two classes.

We conduct these comparisons on the full sample (as opposed to the sibling sub-sample), for two reasons. First, it makes sense to look at gender differences using the full sample, since limiting the sample to sisters and brothers would necessitate small samples and would introduce serious sample size limitations. Second, because of the fact that siblings are separated by age and likely by class, a comparison across class is not possible in the siblings sub-sample.

Before running the regressions, we re-center the dependent variable, so that the normalized score corresponds to the individual group being analyzed, and we then are able to compare like with like. When we split the full sample along gender lines, the results are very similar. In Tables 14 and 15, the point estimates are extremely close, suggesting that there is no discernable difference across genders. Separate results for classes 2 and 4 are presented in Tables 16 and 17. These results are broadly consistent with the results in Table 5, though it does appear that HAZ has a stronger relationship with both Hindi and math for the Class 2 children. In addition, HAZ is not statistically significant in Class 4 school fixed effects regression for Hindi (Column [3], Table 17). However, the coefficients for math are relatively close. Though these results are interesting, there does not seem to be a compelling reason to break down the sample in the analysis that follows.

Table 14: Girls Only, Selected Regressors

	Hindi			Math		
	OLS	Village FE	School FE	OLS	Village FE	School FE
	(1)	(2)	(3)	(4)	(5)	(6)
HAZ	.049	.041	.038	.075	.071	.071
	(.014)***	(.015)***	(.015)**	(.014)***	(.015)***	(.015)***
Class 4	.762	.927	.899	.669	.760	.737
	(.078)***	(.072)***	(.073)***	(.065)***	(.065)***	(.065)***
Age (years)	.036	.021	.016	.065	.052	.053
	(.018)**	(.018)	(.019)	(.017)***	(.017)***	(.017)***
Days of illness (last 3 mos)	-.005	-.005	-.005	-.005	-.005	-.004
	(.001)***	(.001)***	(.001)***	(.002)***	(.002)***	(.001)***
Absent 3+ days (last 2 wks)	-.158	-.135	-.142	-.162	-.146	-.157
	(.037)***	(.037)***	(.036)***	(.036)***	(.035)***	(.035)***
Wave 2	.220	.220	.224	.279	.258	.271
	(.027)***	(.030)***	(.023)***	(.027)***	(.028)***	(.023)***
Mother's education (years)	.024	.018	.015	.021	.017	.014
	(.008)***	(.008)**	(.008)*	(.008)***	(.008)**	(.008)*
Father's education (years)	.015	.018	.017	.013	.012	.011
	(.004)***	(.004)***	(.004)***	(.004)***	(.004)***	(.004)***
Private school	.340	.437		.143	-.037	
	(.234)	(.229)*		(.207)	(.206)	
Obs.	3636	3636	3636	3636	3636	3636
Groups		129	154		129	154
R^2	.424	.396	.356	.452	.386	.354

The dependent variable is normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

Table 15: Boys Only, Selected Regressors

	Hindi			Math		
	OLS	Village FE	School FE	OLS	Village FE	School FE
	(1)	(2)	(3)	(4)	(5)	(6)
HAZ	.050	.034	.021	.086	.073	.064
	(.013)***	(.013)**	(.013)*	(.012)***	(.013)***	(.013)***
Class 4	.703	.728	.775	.684	.704	.723
	(.061)***	(.064)***	(.063)***	(.051)***	(.054)***	(.056)***
Age (years)	.074	.071	.049	.122	.117	.104
	(.016)***	(.016)***	(.016)***	(.014)***	(.014)***	(.015)***
Days of illness (last 3 mos)	-.004	-.003	-.002	-.003	-.003	-.003
	(.001)***	(.001)**	(.001)**	(.001)***	(.001)***	(.001)***
Absent 3+ days (last 2 wks)	-.158	-.151	-.152	-.131	-.125	-.122
	(.034)***	(.031)***	(.031)***	(.033)***	(.031)***	(.030)***
Wave 2	.229	.247	.254	.298	.292	.302
	(.028)***	(.030)***	(.024)***	(.026)***	(.026)***	(.022)***
Mother's education (years)	.006	.008	.009	.013	.016	.016
	(.007)	(.007)	(.006)	(.007)*	(.008)**	(.007)**
Father's education (years)	.020	.018	.017	.017	.015	.014
	(.005)***	(.005)***	(.005)***	(.005)***	(.004)***	(.004)***
Private school	.344	.650		.520	.645	
	(.144)**	(.222)***		(.127)***	(.180)***	
Obs.	4206	4206	4206	4206	4206	4206
Groups		133	158		133	158
R^2	.425	.365	.313	.508	.424	.38

The dependent variable is normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

Table 16: Class 2 Children, Selected Regressors

	Hindi			Math		
	OLS	Village FE	School FE	OLS	Village FE	School FE
	(1)	(2)	(3)	(4)	(5)	(6)
HAZ	.084	.057	.050	.117	.088	.082
	(.014)***	(.013)***	(.013)***	(.011)***	(.013)***	(.012)***
Girl	-.199	-.204	-.203	-.287	-.305	-.307
	(.044)***	(.041)***	(.041)***	(.039)***	(.040)***	(.039)***
Age (years)	.066	.063	.046	.120	.107	.093
	(.019)***	(.018)***	(.018)***	(.018)***	(.017)***	(.017)***
Days of illness (last 3 mos)	-.002	-.002	-.001	-.001	-.002	-.002
	(.001)	(.001)	(.001)	(.001)	(.001)*	(.001)
Absent 3+ days (last 2 wks)	-.127	-.121	-.126	-.115	-.114	-.114
	(.041)***	(.039)***	(.038)***	(.037)***	(.037)***	(.036)***
Wave 2	.311	.318	.336	.382	.384	.396
	(.034)***	(.044)***	(.031)***	(.036)***	(.038)***	(.031)***
Mother's education (years)	.022	.022	.020	.016	.015	.014
	(.009)**	(.009)**	(.009)**	(.009)*	(.009)*	(.009)
Father's education (years)	.018	.016	.016	.015	.015	.014
	(.005)***	(.005)***	(.005)***	(.005)***	(.005)***	(.005)***
Private school	.055	1.707		.250	1.124	
	(.193)	(.757)**		(.191)	(.374)***	
Obs.	3894	3894	3894	3894	3894	3894
Groups		131	156		131	156
R^2	.323	.261	.13	.407	.294	.182

The dependent variable is normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

Table 17: Class 4 Children, Selected Regressors

	Hindi			Math		
	OLS	Village FE	School FE	OLS	Village FE	School FE
	(1)	(2)	(3)	(4)	(5)	(6)
HAZ	.036	.020	.019	.065	.058	.057
	(.017)**	(.016)	(.017)	(.016)***	(.015)***	(.016)***
Girl	-.142	-.119	-.121	-.377	-.360	-.355
	(.037)***	(.037)***	(.037)***	(.037)***	(.038)***	(.038)***
Age (years)	.053	.033	.031	.087	.077	.076
	(.018)***	(.018)*	(.018)*	(.016)***	(.016)***	(.016)***
Days of illness (last 3 mos)	-.006	-.005	-.005	-.006	-.005	-.005
	(.001)***	(.001)***	(.001)***	(.001)***	(.001)***	(.001)***
Absent 3+ days (last 2 wks)	-.224	-.190	-.193	-.179	-.168	-.166
	(.044)***	(.039)***	(.040)***	(.039)***	(.032)***	(.033)***
Wave 2	.218	.224	.230	.266	.245	.265
	(.032)***	(.028)***	(.025)***	(.031)***	(.028)***	(.023)***
Mother's education (years)	.014	.013	.013	.019	.017	.015
	(.007)**	(.007)*	(.007)*	(.008)**	(.008)**	(.008)*
Father's education (years)	.023	.020	.020	.019	.014	.014
	(.004)***	(.005)***	(.005)***	(.004)***	(.004)***	(.004)***
Private school	.506	-.483		.564	.374	
	(.169)***	(.518)		(.160)***	(.251)	
Obs.	3948	3948	3948	3948	3948	3948
Groups		132	157		132	157
R^2	.332	.224	.12	.412	.238	.164

The dependent variable is normalized to have a mean of 0 and standard deviation of 1. Standard errors in parentheses; all standard errors are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

D First stage results

Table 18: First Stage Results for Linear Models, Coefficients on Instruments

Dependent Variable	HAZ	
	2SLS	HH FE
	(1)	(2)
jan-rain-birthyear	.010 (.015)	.012 (.015)
jan-rain-age1	-.005 (.031)	-.015 (.029)
jan-rain-age2	.014 (.025)	.016 (.028)
feb-rain-birthyear	-.015 (.015)	-.016 (.016)
feb-rain-age1	-.010 (.028)	.004 (.026)
feb-rain-age2	.0004 (.024)	.024 (.025)
mar-rain-birthyear	-.011 (.034)	-.020 (.031)
mar-rain-age1	.007 (.040)	-.006 (.034)
mar-rain-age2	-.014 (.037)	-.027 (.030)
apr-rain-birthyear	.033 (.035)	.074 (.036)**
apr-rain-age1	.062 (.041)	.082 (.040)**
apr-rain-age2	.060 (.033)*	.070 (.028)**
may-rain-birthyear	.008 (.008)	.004 (.009)
may-rain-age1	-.008 (.008)	-.008 (.009)
may-rain-age2	.007 (.009)	.008 (.010)
jun-rain-birthyear	-.0002 (.004)	.001 (.004)
jun-rain-age1	.003 (.003)	.003 (.003)
jun-rain-age2	.005 (.004)	-.0001 (.004)
jul-rain-birthyear	.002 (.003)	.005 (.003)*
jul-rain-age1	.0006 (.002)	-.002 (.002)
jul-rain-age2	.006 (.002)***	.004 (.002)**
aug-rain-birthyear	.002 (.003)	.0006 (.003)
aug-rain-age1	.001 (.002)	.0006 (.002)
aug-rain-age2	-.0002 (.004)	-.003 (.004)
sep-rain-birthyear	.0008 (.003)	.005 (.002)**
sep-rain-age1	.002 (.005)	.003 (.004)
sep-rain-age2	.005 (.003)	.002 (.003)
oct-rain-birthyear	.005 (.006)	.005 (.006)
oct-rain-age1	.006 (.005)	.007 (.004)
oct-rain-age2	.003 (.003)	.002 (.003)
nov-rain-birthyear	-.0008 (.013)	.017 (.014)
nov-rain-age1	.005 (.012)	-.013 (.013)
nov-rain-age2	-.011 (.024)	.009 (.020)
dec-rain-birthyear	-.041 (.018)**	-.040 (.017)**
dec-rain-age1	.019 (.028)	.030 (.027)
dec-rain-age2	.094 (.029)***	.025 (.027)
Obs.	1436	1436
F statistic	79.978	37.632
R^2	.325	.427

Standard errors in parentheses are robust to heteroskedasticity and are clustered at the village level. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively.

E Zero-inflated and hurdle models

We can improve upon the efficiency of our estimates by pursuing alternative parametric count data models. In particular, because of the large number of zeros seen in the distribution in Figures 2 and 3, a zero-inflated model may be appropriate. A zero-inflated model adjusts the likelihood function so that there are two processes: one that creates a zero count, and one that creates a zero or a positive count. As with other count models, one can assume an underlying distribution for the data. In Tables 19 and 20, we show results for the zero-inflated Poisson and negative binomial estimators.

Table 19: Zero-Inflated Counts and Hurdle Models for Hindi, Selected Regressors

	Zero-Inflated Poisson	Hurdle Poisson	Hurdle NegBin I
	(1)	(2)	(3)
Positive counts: $Total\,score > 0$			
HAZ	.085	.087	.134
	(.025)***	(.025)***	(.026)***
Inflate/Hurdle: $\Pr(Total\,score = 0)$			
HAZ	-.365	-.365	-.365
	(.136)***	(.136)***	(.136)***
Obs.	1436	1436	1436
No. of Zero Obs.	87	87	87
Log Likelihood	-21778.36	-21778.36	-6553.536

Standard errors in parentheses are robust to heteroskedasticity. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively. Columns [2] and [3] use the Stata commands `hplogit` and `hnblogit`, respectively.

Table 20: Zero-Inflated Counts and Hurdle Models for Math, Selected Regressors

	Zero-Inflated Poisson	Hurdle Poisson	Hurdle NegBin I
	(1)	(2)	(3)
Positive counts: $Total\,score > 0$			
HAZ	.087	.087	.115
	(.023)***	(.023)***	(.024)***
Inflate/Hurdle: $\Pr(Total\,score = 0)$			
HAZ	-.585	-.585	-.585
	(.161)***	(.161)***	(.161)***
Obs.	1436	1436	1436
Number of zeros	78	78	78
Log likelihood	-16739.46	-16739.46	-6673.141

Standard errors in parentheses are robust to heteroskedasticity. Regressions include several other controls—see Table 13 in Appendix B for a full set of RHS variables. Stars (*, **, ***) indicate levels of significance of 90%, 95% and 99%, respectively. Columns [2] and [3] use the Stata commands `hplogit` and `hnblogit`, respectively.

A final estimation strategy that we pursue is to consider more closely what we think is driving the test score process. If we believe that the process that drives scores of zero is somehow fundamentally different from the process that drives positive scores, then a hurdle model is a good alternative to a zero-inflated count regression.[35] A hurdle model is in spirit like a Heckman selection approach—a logit is used to determine whether one has a positive score or a zero, and then a zero-truncated count regression (one that allows the distribution, either Poisson or NegBin I, to lack zeros) is run on the positive counts. The flexibility inherent in this approach is intuitively appealing.

[35] Following Cameron and Trivedi (1999), the zeros and the positive counts are determined by two different densities, $f_1(.)$ and $f_2(.)$, such that $\Pr[y=0] = f_1(0)$ and $f(y|y>0) = \frac{f_2(y)}{1-f_2(0)}$. Logically, $\Pr[y>0] = 1 - f_1(0)$. It follows that data generating process is:

$$g(x) = \begin{cases} f_1(0) & \text{if } y = 0, \\ \frac{1-f_1(0)}{1-f_2(0)} f_2(y) & \text{if } y \geq 0. \end{cases}$$

References

Alderman, H., Behrman, J., Lavy, V., & Menon, R. (2001). Child health and school enrollment: A longitudinal analysis. *Journal of Human Resources*, *36*(1), 185-205.

Alderman, H., Hoddinott, J., & Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers*, *58*, 450-474.

Allison, P., & Waterman, R. (2002). Fixed-effects negative binomial regression models. *Sociological Methodology*, *32*, 247-265.

Averett, S., & Stifel, D. (2007). *Food for thought: The cognitive effects of childhood malnutrition in the United States.* (Unpublished Working Paper)

Banerji, R., & Kingdon, G. (2009). *Understanding teacher competencies in north India.* (Presentation to the 'Teacher Development and Management Conference', Udaipur, 23-25 February)

Beegle, K., De Weerdt, J., & Dercon, S. (2006). Orphanhood and the long-run impact on children. *American Journal of Agricultural Economics*, *88*(5), 12661272.

Behrman, J. (1996). The impact of health and nutrition on education. *The World Bank Observer*, *11*(1), 23-37.

Behrman, J., & Lavy, V. (1997). *Child health and schooling achievement: Association, causality, and household allocations.* (Penn Institute for Economic Research Working Paper 97-023)

Bobonis, G., Miguel, E., & Puri-Sharma, C. (2006). Anemia and school participation. *Journal of Human Resources*, *XLI*(4), 692-721.

Borooah, V. (2005). The height-for-age of Indian children. *Economics and Human Biology*, *3*(1), 45-65.

Bound, J., Jaeger, D., & Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association*, *90*, 443-450.

Brennan, L., McDonald, J., & Shlomowitz, R. (2004). Infant feeding practices and chronic child malnutrition in the Indian states of Karnataka and Uttar Pradesh. *Economics and Human Biology*, *2*(1), 139-158.

Cameron, A., & Trivedi, P. (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*, 29-53.

Cameron, A., & Trivedi, P. (1999). *Essentials of count data regression.* (Available at http://cameron.econ.ucdavis.edu/research/CTE01preprint.pdf)

Citizens Initiative for the Rights of Children Under Six. (2006, December). *Focus on children under six.* Available from `unicef.globalstudy.googlepages.com/FocusonChildrenUnderSixDecember2006.pdf` (UNICEF, Abridged Report)

Cole, T., Flegal, K., Nicholls, D., & Jackson, A. (2007). Body mass index cut offs to define thinness in children and adolescents: International survey. *British Medical Journal*, *335*(194), 1-8.

Datar, A., & Sturm, R. (2006). Childhood overweight and elementary school outcomes. *International Journal of Obesity*, *30*, 1449-1460.

de Onis, M., Garza, C., Onyango, A., & Borghi, E. (2007). Comparison of the WHO child growth standards and the CDC 2000 growth charts. *Journal of Nutrition*, *137*, 144-148.

Deaton, A. (2008). Height, health, and inequality: the distribution of adult heights in India. *American Economic Review*, *98*(2), 468474.

Deaton, A., & Drèze, J. (2008). *Nutrition in India: Facts and interpretations.* Available from `http://weblamp.princeton.edu/chw/papers/ deaton_dreze_india_nutrition.pdf`

Dercon, S., & Hoddinott, J. (2003). Health, shocks, and poverty persistence. In S. Dercon (Ed.), *Insurance against poverty.* Oxford: Oxford University Press.

Dercon, S., & Krishnan, P. (2000). In sickness and in health: Risk sharing within households in rural Ethiopia. *Journal of Political Economy*, *108*(4), 688-727.

Dobbing, J. (1976). Vulnerable periods in brain growth and somatic growth. In D. Roberts & A. Thomson (Eds.), *The biology of human fetal growth.* London: Taylor and Francis.

Drèze, J., & Kingdon, G. (2001). School participation in rural India. *Review of Development Economics*, *5*(1), 1-24.

Flegal, K., Wei, R., & Ogden, C. (2002). Weight-for-stature compared with body mass indexfor-age growth charts for the United States from the Centers for Disease Control and Prevention. *American Journal of Clinical Nutrition*, *75*, 761-766.

Gibson, J. (2001). Literacy and intrahousehold externalities. *World Development*, *29*(1), 155-166.

Glewwe, P. (1999). Why does mother's schooling raise child health in developing countries? Evidence from Morocco. *Journal of Human Resources*, *34*(1), 124-159.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature*, *40*(2), 436-482.

Glewwe, P. (2005). The impact of child health and nutrition on education in developing countries: Theory, econometric issues, and recent empirical evidence. *Food and Nutrition Bulletin*, *26*(2 (supplement 2)), S235-S250.

Glewwe, P., & Jacoby, H. (1995). An economic analysis of delayed primary school enrollment and childhood malnutrition in a low income country. *The Review of Economics and Statistics*, *77*(1), 156-169.

Glewwe, P., Jacoby, H., & King, E. (2001). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics*, *81*, 345-368.

Glewwe, P., & King, E. (2001). The impact of early childhood nutritional status on cognitive development: Does the timing of malnutrition matter? *World Bank Economic Review*, *15*(1), 81-113.

Glewwe, P., & Miguel, E. (2008). The impact of child health and nutrition on education in less developed countries. In T. Schultz & J. Strauss (Eds.), *Handbook of development economics volume 4* (p. 3561-3606). Amsterdam: Elsevier Science.

Gomes-Neto, J., Hanushek, E., Leite, R., & Frota-Bezzera, R. (1997). Health and schooling: Implications for developing countries. *Economics of Education Review*, *16*(3), 271-282.

Hoddinott, J., & Kinsey, B. (2001). Child growth in the time of drought. *Oxford Bulletin of Economics and Statistics*, *63*(4), 409-436.

Jalan, J., & Ravallion, M. (2003). Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics*, *112*(1), 153 - 173.

Kaestner, R., & Grossman, M. (2008). *Effects of weight on children's educational attainment.* (NBER Working Paper No. 13764)

Kingdon, G. (2007). The progress of school education in India. *Oxford Review of Economic Policy*, *23*(2), 168-195.

Kremer, M., Chaudhury, N., Rogers, F., Muralidharan, K., & Hammer, J. (2005). Teacher absence in India: A snapshot. *Journal of the European Economic Association*, *3*(2-3), 658-667.

Lavy, V., Strauss, J., Thomas, D., & de Vreyer, P. (1996). Quality of health care, survival and health outcomes in Ghana. *Journal of Health Economics*, *15*, 333-357.

Lee, L. fei, Rosenzweig, M. R., & Pitt, M. M. (1997). The effects of improved nutrition, sanitation, and water quality on child health in high-mortality populations. *Journal of Econometrics*, *77*(1), 209 - 235.

Martorell, R. (1995). Results and implications of the INCAP follow-up study. *Journal of Nutrition*, *125*(4), 1127S-1138S.

Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*(1), 159-217.

Mullahy, J. (1997). Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *The Review of Economics and Statistics*, *79*(4), 586-593.

Muralidharan, K., & Kremer, M. (2006). Public and private schools in rural India. In R. Chakrabarti & P. Peterson (Eds.), *School choice international: Exploring public-private partnerships.* Cambridge: MIT Press.

Nichols, A. (2007). *IVPOIS: Stata module to estimate an instrumental variables poisson regression via GMM.* Available from `http://ideas.repec.org/c/boc/bocode/s456890.html`

O'Donnell, O., Doorslaer, E. van, Wagstaff, A., & Lindelow, M. (2007). *Analyzing health equity using household survey data: A guide to techniques and their implementation.* Washington, DC: The World Bank.

Ogden, C., Kuczmarski, R., Flegal, K., Mei, Z., Guo, S., Wei, R., et al. (2002). Centers for Disease Control and Prevention 2000 Growth Charts for the United States: Improvements to the 1977 National Center for Health Statistics Version. *Pediatrics*, *109*(1), 45-60.

Pal, S., & Kingdon, G. (2009). *Private school growth and universal literacy in India: A district-level analysis, 1992-2002.* (Mimeo, Brunel University)

Pollitt, E., Gorman, K., Engle, P. L., Rivera, J. A., & Martorell, R. (1995). Nutrition in early life and the fulfillment of intellectual potential. *Journal of Nutrition*, *125*(4), 1111S-1118S.

Pratham. (2008, January). *Annual status of education report 2007.* New Delhi.

Pratham. (2009, January). *Annual status of education report 2008.* New Delhi.

Rosenzweig, M., & Wolpin, K. (1986). Evaluating the effects of optimally distributed public programs: Child health and family planning interventions. *American Economic Review*, *76*(3), 470-482.

Rosenzweig, M., & Wolpin, K. (1994). Are there increasing returns to the intergenerational production of human capital? Maternal schooling and child intellectual achievement. *Journal of Human Resources*, *29*(2), 670-693.

Sabia, J. (2007). The effect of body weight on adolescent academic performance. *Southern Economic Journal*, *73*(4), 871-900.

Sandiford, P., Cassel, J., Montenegro, M., & Sanchez, G. (1995). The impact of women's literacy on child health and its interaction with health services. *Population Studies*, *49*(1), 5-17.

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. Andrews & J. Stock (Eds.), *In identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (p. 80-108). Cambridge: Cambridge University Press.

Tan, J.-P., Lane, J., & Lassibille, G. (1999). Outcomes in Philippine elementary schools: An evaluation of four experiments. *World Bank Economic Review*, *13*(3), 493508.

Tanner, J., Healy, M., Lockhart, R., MacKenzie, J., & Whitehouse, R. (1956). Aberdeen growth study. i. the prediction of adult body measurements from measurements taken each year from birth to 5 years. *Archives of Disease in Childhood*, *31*(159), 372381.

Tarozzi, A. (2008). Growth reference charts and the nutritional status of Indian children. *Economics and Human Biology*, *6*(3), 455-468.

Tarozzi, A., & Mahajan, A. (2007). Child nutrition in India in the nineties. *Economic Development and Cultural Change*, *55*, 441486.

Thomas, D., Lavy, V., & Strauss, J. (1991). Public policy and anthropometric outcomes in the Côte d'Ivoire. *Journal of Public Economics*, *61*(2), 155-192.

Thomas, D., Strauss, J., & Henriques, M.-H. (1991). How does mother's education affect child height? *Journal of Human Resources*, *26*(1), 183-211.

UNESCO. (2006). *Education for All: Global monitoring report 2007.* Available from http://unesdoc.unesco.org/images/0014/001416/141639e.pdf

Waber, D., Vuori-Christiansen, L., Ortiz, N., Clement, J., Christiansen, N., Mora, J., et al. (1981). Nutritional supplementation, maternal education, and cognitive development of infants at risk of malnutrition. *American Journal of Clinical Nutrition*, *34*, 807-813.

Wisniewski, S. (2009). Child nutrition, health problems, and school achievement in Sri Lanka. *World Development*. (forthcoming, doi:10.1016/j.worlddev.2009.09.009.)

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.

Wu, K. B. (2009). International benchmarking and determinants of mathematics achievement in two Indian states. *Education Economics*. (forthcoming)