On the Demand for Education in India

On the Demand for Education in India

A dissertation presented

by

Mary Bryce Millett Steinberg

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

April 2015

Dissertation Advisor:
Professor Michael Kremer

Author:
Mary Bryce Millett Steinberg

On the Demand for Education in India

Abstract

In this dissertation I examine the impacts of market forces and government programs on households' demand for human capital in India. The first chapter examines the impact of ITES Centers on school enrollment using administrative enrollment data from three states in India, and finds that when these centers open, enrollment in primary school increases significantly. The effects are very localized, and using supplementary survey evidence we argue that this is driven by limited information diffusion. The second chapter introduces a simple model of human capital production which predicts that wages can negatively impact human capital under reasonable assumptions. Using data on test scores and schooling from rural India, we show that human capital investment is procyclical in early life (in utero to age 3) but then becomes countercyclical. We argue that, consistent with our model, this countercyclical effect is caused by families investing more time in schooling when outside options are worse. The final chapter applies the findings from this study to understand how workfare programs (a common anti-poverty strategy in the developing world) can impact school enrollment through their effects on wages. We examine the effect of the largest anti-poverty workfare program in world: NREGA in India. Using a fixed effects estimator, I show that the introduction of NREGA caused increases in child employment, and decreases in school enrollment, particularly among children ages 13-17.

Contents

List of Tables

List of Figures

Acknowledgments

I am extremely grateful to have had the constant advice and support of my colleagues, mentors, family and friends throughout graduate school.

I am deeply indebted to my two co-authors, Emily Oster and Manisha Shah, without whom this research would not have been possible (or at least would not have been nearly as much fun). They are both extremely insightful empirical researchers, and working with them has taught me a tremendous amount about how to ask interesting questions and answer them convincingly. I am incredibly grateful for their friendship, advice, and patience.

I owe a tremendous amount to my advisors for their constant encouragement and support. Ed Glaeser has always pushed me to be better, clearer, and more insightful. Nava Ashraf has taught me the importance of both theory and context in my research. Michael Kremer has encouraged me to think bigger, and to work tirelessly toward excellence. In addition, I am grateful to my undergraduate advisor, David Weil, without whom it is extremely unlikely I would have considered a career in research. I'm really glad I did.

At various stages, these papers benefitted enormously from the helpful feedback of many colleagues, including Natalie Bau, Thomas Covert, David Cutler, James Feigenbaum, Roland Fryer, Peter Ganong, Matt Gentzkow, Benjamin Hebert, Nathan Hendren, Rick Hornbeck, Simon Jaeger, Seema Jayachandran, Robert Jensen, Larry Katz, Benjamin Lockwood, Matt Notowidigdo, Guillaume Pouliot, Jesse Shapiro, Andrei Shleifer, and Martin Rotemberg. I am also extremely appreciative for the institutional support from Karla Cohen, Dianne Le, Charlotte Tuminelli, Jeanne Burke, Jeanne Winner, and Brenda Piquet, and to Nir Hak, Xiaoman Luo, and Perwinder Singh for excellent research assistance.

Finally, I am grateful to my parents, Kathleen Scott, Joseph Millett, Merrill Scott, and Suzanne Halpin, my grandmother Paige Tourville, and my brother Michael Millett, for their love, support, and encouragement. And of course, to my husband, Jeff Steinberg, who makes the whole thing worthwhile.

To my parents, and to J

Introduction

In much of the developing world, school enrollment rates and achievement remain low despite large investments in lowering the costs of schooling to households. In this dissertation I examine the impacts of market forces and government programs on households' demand for human capital in India.

In the first chapter, I use microdata to evaluate the impact of ITES (call center) jobs on local school enrollment in areas outside of major IT centers in southern India. I merge panel data on school enrollment from a comprehensive school-level administrative dataset with detailed data on Information Technology Enabled Services (ITES) center location and founding dates. Using school fixed effects, I find that introducing a new ITES center causes a 5% increase in the number of children enrolled in primary school; this effect is localized to within a few kilometers. I show the effect is driven by English-language schools, consistent with the claim that the impacts are due to changes in returns to schooling, and is not driven by changes in population or income resulting from the ITES center. Supplementary survey evidence suggests that the localization of the effects is driven by limited information diffusion.

In the second and third chapters, I in turn examine the effects of opportunity costs of schooling for households, in the form of higher wages. Higher wages are generally thought to increase human capital production, particularly in the developing world. However, in the second chapter, I introduce a simple model of human capital production which predicts that wages can negatively impact human capital under reasonable assumptions. Using data on test scores and schooling from rural India, I show that human capital investment is procyclical in early life (in utero to age 3) but then becomes countercyclical. I argue that, consistent with our model, this countercyclical effect is caused by families investing more time in schooling when outside options are worse. In addition, we find long term impacts of these shocks: adults who experienced more positive rainfall shocks during school years have lower overall total years of schooling. These results suggest that the opportunity cost of schooling, even for fairly young children, is an important factor in determining overall human capital investment.

In the final chapter, I apply the findings from this study to understand how workfare programs (a common anti-poverty strategy in the developing world) can impact school enrollment through their effects on wages. I examine the effect of the largest anti-poverty workfare program in world: NREGA in India. Using a fixed effects estimator, I show that the introduction of NREGA caused increases in child employment, and decreases in school enrollment, particularly among children ages 13-17. In addition, math and reading test scores and grade levels are decreased at the introduction of the program. We show that while effects are similar for boys and girls, boys are substituting more into outside work, while girls are substituting for their mothers inside the home. We conclude that anti-poverty programs which raise wages could have the perverse effect of lowering human capital investment in the developing world.

This research shows that households respond to both changes in perceived costs and benefits of schooling, and that liquidity does not seem to explain low overall rates of school enrollment. Instead, it is possible that information barriers, heterogeneous returns or perhaps intergenerational contracting problems could do more to explain the low levels of human capital in India. In addition, I argue that because opportunity costs and perceived benefits matter for households' schooling decisions, and thus government programs with the potential to change prices should be aware of their potential effects on human capital, and subsequent economic growth.

Chapter 1

Do ITES Centers Improve School Enrollment? Evidence from India [1]

1.1 Introduction

In the last thirty years, globalization has dramatically changed job opportunities in the developing world. In many countries this change has increased the skill premium. In India, the focus of this paper, this change has been particularly striking. The number of individuals employed in outsourcing-related businesses has increased from roughly 50,000 in 1991 to over 2 million in 2010 (NASSCOM, 2010); these jobs demand employees with high levels of education and a good command of English, and pay high salaries by Indian standards. The availability of these new opportunities increases the return to education which may, in turn, increase school enrollment, a possibility often floated in popular media (i.e. Giridharadas, 2010).[2] Although India has made significant strides in schooling, even primary school completion is not universal. In the 2005-2006 National Family and Health Survey, for example, only 79.2% of children aged 6 to 14 were enrolled in school.

[1] Co-authored with Emily Oster

[2] This question echoes a very large existing literature on the returns to education and school enrollment in both the developing and developed world (e.g. Freeman, 1976; Katz and Murphy, 1992; Heckman, 1993; Kane, 1994; Foster and Rosenzweig, 1996; Griliches, 1997).

An academic literature suggests that cities and districts with a major IT presence have experienced changes in education patterns with the growth in these jobs (Munshi and Rosenzweig, 2006; Shastry, 2012). In a randomized evaluation, Jensen (2012) shows that targeted recruitment for these jobs in areas outside major cities also influences schooling choices. Together, these papers suggest that increasing the presence of ITES centers in new areas *could* increase school enrollment in those areas. Without the presence of targeted information campaigns, will the existence of ITES centers in the local area cause parents to change their investment in their children's human capital?

It is this question we address in this paper. To do so, we first estimate the effect of the introduction of these businesses on school enrollment outside of the major IT areas. This allows us to evaluate the validity of the popular claim that these businesses will have broad geographic impacts in all of India. Our data is sufficient to allow us to distinguish the magnitude of impacts over quite small distances, and we argue we are able to make strong causal statements. As we detail below, we find that the impacts of IT centers on school enrollment are large but localized.

Second, with a more qualitative survey we are able to provide some preliminary evidence on the mechanisms behind these effects and their relatively narrow geographic range. We argue this effect is due to limited information dissemination across areas.[3] This suggests that in the absence of any intervention impacts may *not* be geographically broad, although better information provision about job opportunities could have large impacts.

We begin by using panel data on schooling and the introduction of Information Technology Enabled Services (ITES) firms[4] to estimate the impact of these businesses on school enrollment. Our school enrollment data come from a comprehensive administrative dataset and cover three states in India (Karnataka, Andhra Pradesh and Tamil Nadu); each school is observed for a period of four to eight years between 2001 and 2008. We combine this with a

[3]This is consistent with Jensen (2010) and Jensen (2012), both of which suggest interventions which provide better information on job opportunities (in the former case, in a very similar setting) change schooling decisions.

[4]This is a class of business which includes call centers, as well as data processing, medical imaging and related services.

newly collected dataset on ITES business locations and founding dates. Our ITES center data includes areas outside of Chennai, Hyderabad and Bangalore, which allows us to estimate the impact of jobs in areas which have not had an overwhelming IT presence. Our ITES center location data allows us to identify the PIN code (similar to a ZIP code) location of each center, which we can link to school location. We use a school fixed effects estimator to analyze how enrollment changes *within an individual school* upon the introduction of a new ITES center to the area.

We estimate the impact of ITES center introduction on schools in the same PIN code and find strong positive effects: the introduction of one additional ITES center to the PIN code is associated with an increase between 4% and 7% in number of children enrolled in the school in the year after the center introduction.[5] In addition to school fixed effects, we control for several time-varying school infrastructure controls and year fixed effects interacted with state dummies and area demographics. Our preferred specification is one in which we limit to areas most comparable to the areas with ITES centers: areas with English-language schools. This specification yields a coefficient of 4.8%. Our effects are robust to controlling for district specific trends and to limiting to areas which ever have ITES centers.

We next estimate the geographic range of these impacts, and find they are very localized. ITES centers in PIN codes 5 kilometers or less away have a significant positive impact on school enrollment, but it is very small. ITES centers in PIN codes between 5 and 10 kilometers away have no impact.

An issue with interpreting these results causally is the possibility that the introduction of ITES centers anticipates increased school enrollment rather than causing it. The inclusion of school fixed effects in our specification addresses the concern that ITES center introduction is associated with some fixed area characteristic, but they do not address the concern that ITES centers might be introduced to areas which are *changing* more rapidly.[6] To address

[5]This effect is driven in large part by older children, which is consistent with the fast impact.

[6]We should note that we have no reason to think this type of endogenous placement is common. Conversations with ITES center operators suggested they choose where to locate primarily based on the level of infrastructure and the quality of possible employees, but there was no mention of locating based on anticipated

this issue, we analyze the impact of ITES centers introduced in future years on current enrollment.[7] If ITES center operators are targeting areas which have more rapidly increasing schooling, future ITES centers should also correlate with changes in schooling. Similarly, if other variables are changing continuously and driving both variables, then future ITES centers should correlate with current changes. We do not find evidence for an impact of future ITES centers and the inclusion of the future ITES center measure does not affect our estimate of the impact of current ITES centers.

We also explore whether these impacts vary by language of instruction. ITES jobs almost universally require knowledge of English in addition to high levels of education. Consistent with this, we find that enrollment in English-language schools increases by around 7% with the introduction of each ITES center, whereas there is no significant change for local-language schools.

In terms of magnitude, the results suggest that introducing an ITES center (median size of 80 employees) increases enrollment in the PIN code by 180 children. We can also describe this in terms of share of out-of-school children who enroll. In the three states we use, the nationally representative National Family and Health Survey shows 84.4% of children in this age group reported attending school at all during the 2005-2006 school year. Using our preferred coefficient of 4.8%, this suggests that about 26% of out-of-school children are enrolled as a result of ITES center introduction.

These results point to a causal impact of ITES centers on school enrollment. In robustness checks we address several lingering concerns – that the results reflect changes in population after an ITES center opening, that they reflect changes in income and that they reflect changes in the number of schools – and argue these issues do not drive our results.

In Section 6 we explore mechanisms. We distinguish two possibilities. First, the introduction of an ITES center may impact actual returns to schooling in the local area by

increases in schooling or previous years schooling increases.

[7] This methodology has been used elsewhere to test for similar concerns (Jensen and Oster, 2009; La Ferrara, Chong and Duryea, 2009).

providing new jobs at *that* center; this can explain our results only if labor markets are very localized. Alternatively, it may impact perceived returns to schooling by providing better information about these jobs in general; this can explain our results only if information is very localized.

To distinguish between these possible mechanisms, we conducted a survey in one district in Tamil Nadu (Madurai), which allows us to observe (a) the localization of the labor market and (b) how widely information diffuses. We find evidence in favor of the information story. Our data indicate that many people do travel several kilometers for work, which suggests the narrow geographic range of ITES center impacts is not due to localization of labor markets. In contrast, knowledge is very localized. Even limiting the sample to individuals who live *within one kilometer* of an ITES center we find that those who live closer are more likely to report they know of a center in the local area and to correctly identify what qualifications are required for the job.

In addition to the relationship to the literature on IT development and education discussed previously, the findings in this paper relate to a large literature on what policies are effective in increasing school enrollment in the developing world (e.g. Duflo, 2001; Kremer, 2003; Kremer et al., 2005; Duflo, Hanna and Ryan, 2012; Burde and Linden, 2009). In this policy space, our results suggest that better information may be effective in promoting school enrollment in areas further from new job options.[8] We note that although the evidence here focuses on jobs which require high skills, Heath and Mobarak (2012) show evidence that growth in garment factory jobs in Bangladesh also improve schooling for girls, suggesting a similar phenomenon at play even with somewhat lower skill jobs.

The rest of the paper is organized as follows. Section 2 provides some background on ITES centers, and describes the data. Section 3 describes our empirical strategy. Section 4 shows the central results of the paper, and Section 5 discusses robustness. Section 6 presents our survey data and Section 7 concludes.

[8] A caveat to this policy argument is that our results hinge on the fact that jobs in ITES centers require additional education; Atkin (2009) finds that growth in the export sector in Mexico actually leads to school *dropout* since export jobs pay well but do not require schooling.

1.2 Background and Data

1.2.1 Background on ITES Centers

Although the concept of "outsourcing" business processes to low-wage countries has been around since the 1970s, the industry remained small until the late 1990s, as time and cost restrictions were large. With the investment in trans-oceanic fiber-optic cables however, the costs of ITES off-shoring plummeted, and with its relatively educated English-speaking low-wage population, India emerged as the dominant provider of business services ranging from call centers to software development.

ITES center jobs are typically high-paying by Indian standards. The average starting salary at such firms is roughly 8,000 Rupees per month (about US$175), which is almost double the average per capita income of India (Ng and Mitter, 2005). These firms typically come in two types: (1) multinational corporations with subsidiaries or divisions located in India, and (2) Indian "third-party" firms that provide ITES centers and other services for Western companies. Jobs at the Indian firms tend to have lower wages, higher turnover, and less training than the "in-house" multinational corporation positions (Dossani and Kenney, 2004). The majority of ITES centers are in larger cities such as Bangalore, Delhi, and Mumbai, but they are spreading rapidly to smaller cities all over southern India.

Many of these firms are call centers, which focus on direct telephone interaction with Western customers. Workers make outgoing calls (for services like telemarketing), and take incoming calls (for customer service, tech support, and credit card activation, among other things) for large Western companies. At these centers, "voice" workers conduct calls almost entirely in English, primarily to the United States; thus, fluency is generally a requirement for entry-level positions.[9] Other "non-voice" business processes outsourced to such firms range greatly in their skill-level, from data entry to software design. English proficiency may be less important for these jobs, but all of the centers in our study report English as a

[9] Indeed, many of these firms go to great lengths to train their workers to speak with American and British regional dialects, even adopting pseudonyms and memorizing idioms. Some workers report having to watch hours of American television programs to help perfect their speech patterns (Ng and Mitter, 2005).

requirement.

From the perspective of this paper, there are at least two central features of ITES centers which we want to highlight. First, they require relatively high rates of education and pay high wages. To the extent that jobs of this type have not been available historically, their existence may well affect the returns to education (both perceived and actual). Second, the vast majority of these jobs require English skills, which is likely to affect the wage returns to learning English.

1.2.2 Data on School Enrollment

We use a large administrative dataset on primary school enrollment in India called the District Information System for Education (DISE). This dataset has been collected by the Indian government since the late 1990s, although the data used in this paper begins in the early 2000s. Data collection is coordinated at the district level and involves surveys of schools. These school surveys have several parts. First, they collect data on primary school enrollment, including comprehensive data on number of enrolled students by age, grade, gender and caste. These data are designed to reflect enrollment (not attendance) statistics as of September 30th of the school year (which starts in the spring). Second, they collect data on features of the school, including language of instruction and physical plant characteristics. Each school is given a unique ID number, which allows us to follow schools over time.

The area-level survey is less comprehensive and less frequent, but includes some information on village or urban neighborhood characteristics (throughout this paper we will refer to these regions as "neighborhoods"). Most importantly, for most areas in this survey we observe the PIN code location of the school, which allows us to match the area to ITES center locations. A PIN code is similar to a ZIP code in the US; it is smaller than a census block.

The DISE data is collected by the district and then aggregated by each state government. We use data from three states that have been significantly impacted by globalization:

Karnataka, Andhra Pradesh and Tamil Nadu.[10] The number of years of data varies across states. Panel A of Table 1.1 shows, for each state, the years of data coverage and the number of schools in the first and last year. In later years the dataset is more comprehensive, covering a larger share of schools. Although this means we do not have a balanced panel, by including school fixed effects we ensure that we compare the same schools over time. We use all years of the panel. However, the bulk of our changes are in 2005 and 2006 and we will illustrate our results using these groups. In a robustness section we will show results limited to this group.

Panel B of Table 1.1 provides some summary statistics on school enrollment and school characteristics. The average school in our sample is fairly small, with 144 students. The physical plant variables indicate schools are not in very good repair. In an average school, only 70% of classrooms are noted to be in good condition by surveyors. Half of the schools report having a boundary wall, half report having electricity and slightly above half have a toilet. Eleven percent of the sample reports at least some instruction in English (this is based on a question about what languages the school teaches in; they could list as many as they wanted).

This data has several limitations. First, as noted, the coverage of our sample differs somewhat across years. In general, the school fixed effects mean this is not a major issue. The one note of caution is that if the schools we observe are different than the schools we do not observe, our results may have limited generalizablity. This is unlikely to be a serious issue, however, since our best estimates suggest we cover nearly all schools in India.[11]

Second, the DISE data covers only primary schools. It seems plausible, even likely, that much of the impact of ITES centers would be on enrollment in secondary school, since

[10]These three are also states in which we have a relatively long time series of data. Although there are of course many more areas of India, we argue these areas should be representative of areas most heavily impacted by these jobs.

[11]This is actually a somewhat difficult fact to measure. Official statistics on number of schools in India appear to be largely based on the same data we use here so there is no outside source that we can use to verify coverage. The fact that the Indian government uses this as the source of official statistics, however, gives us confidence that we are covering at least an extremely large share of total schools.

Table 1.1: *Summary Statistics*

Panel A: Years of Coverage and Number of Schools			
	Andhra Pradesh	Karnataka	Tamil Nadu
Years of Data Coverage	2004-2007	2001-2007	2003-2007
Number of Schools in:			
First Year of Coverage	59,121	27,136	43,662
Last Year of Coverage (2007)	98,485	52,369	51,548

Panel B: School Summary Statistics			
	Mean	Std. Dev.	Observations
Total Enrollment	143.8	166.4	905,838
% Classrooms in Good Condition	70.7	37.2	905,838
% Schools with Electricity	49.0	50.0	905,838
% Schools with Boundary Walls	51.3	50.0	905,838
Teach in English (0/1)	0.11	0.32	905,717
Total School-Age Population	163.1	1,401	255,355

Panel C: Number of ITES Centers By State		
	Number of ITES Centers	
	Including Cities	Excluding Cities
Andhra Pradesh	100	74
Karnataka	144	121
Tamil Nadu	157	65

Panel D: Number of Schools by Category	
	Number of Schools
Never Had an ITES Center	238,986
Has Same Number of ITES Centers	172
Has Change in Number of ITES Centers	408

Notes: This table shows summary statistics for our sample of schools and ITES centers. Panel A shows years of data coverage and summary statistics by state for the three states in our data set. Panel B shows summary statistics on enrollment and school characteristics for the sample of schools used in the analysis. Population is recorded by the schools for only a subset of schools and years. Panel C shows the number of ITES centers in our sample for each state, including and excluding cities (all later analysis will exclude the major cities of Hyderabad, Chennai, and Bangalore). Location (PIN code) and founding year were collected in a primary survey; only centers with both location and founding date were included in the sample. Panel D shows the number of schools in our analysis which are matched to PIN codes which ever have an ITES center in our data.

secondary school education is typically necessary for these jobs, and enrollment at that level is lower in general. Unfortunately, we cannot observe these enrollments; if anything, this may lead us to understate the impacts.

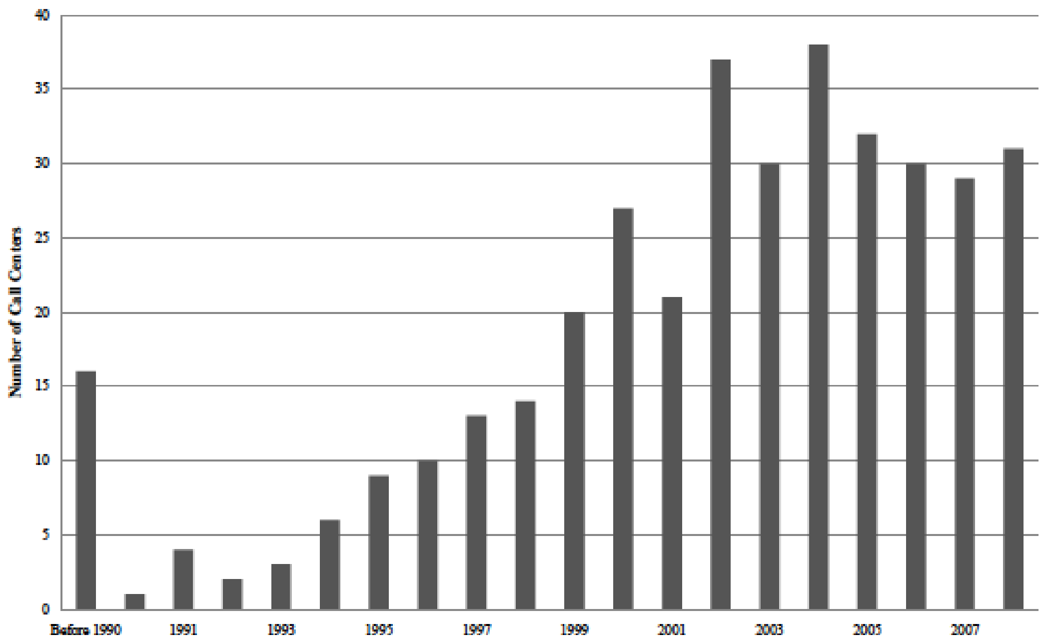
A final important issue is that the data measures total number of children enrolled, not enrollment rates. This leads to the concern that our results reflect changes in population. We discuss this issue in greater detail in Section 5. For a small subset of school years the school also reported the total population of school-aged children in the area. The coverage of these data is limited, but in a robustness check we will use these data and the variable is summarized in Panel B of Table 1.1.

Although we do not observe enrollment rates, we can use the 2005-2006 National Family and Health Survey to estimate the share of children in this age group who are enrolled; this may be important in exploring what effect size is plausible and which groups are likely to be affected. In the states we consider here, enrollment is highest among children in the 8 to 11 year-old age group (around 92%), and lower for kids aged 6 to 7 (74%) and 12 to 14 (80%). This suggests scope for increases in enrollment among both younger children (due to initial enrollment at younger ages) or among older children (likely due to retention).

1.2.3 Data on ITES Centers

To match with the data on education, we collected data on ITES centers. We contracted with a firm in India that helps connect Western firms with Indian ITES centers to create a directory of ITES centers in Andhra Pradesh, Karnataka, and Tamil Nadu. They used their contacts, the Internet, and available directories to compile a list of firms, and called each to confirm their existence, the PIN code of their location and their founding date. Our data collection project focused on areas outside of Bangalore, Chennai and Hyderabad, although we did collect some information on centers there. This focus was in line with our desire to estimate the impacts of these firms outside of major IT centers.

This data collection project resulted in a dataset of 401 ITES centers. Figure 1 shows a histogram of ITES center founding dates; the incredible growth in number of centers over time is clear: in our sample, 68% are founded after the year 2000. As we note above, our data on schooling is collected in September for the year spanning June to April, and the ITES center founding dates are given as simply the calendar year of founding. We code the school year 2005-2006 as 2005, and match with ITES centers this way. A school in a PIN



Notes: This figure shows the distribution of ITES center founding dates among centers in our sample.

Figure 1.1: *Distribution of ITES Center Founding Dates*

code with an ITES center introduced in 2005 is coded as having a new ITES center in the 2005-2006 school year.

The breakdown of number of ITES centers by state is presented in Panel C of Table 1.1. In Column 1 we show the count of all ITES centers; Andhra Pradesh is slightly less well-represented, but the number of ITES centers is fairly similar across states. In Column 2 we report these counts for areas outside the major cities of Bangalore, Chennai and Hyderabad (this is the sample we use for analysis). As expected, this limits the sample somewhat and we are left with 260 ITES centers.

Panel D of Table 1.1 gives a sense of the source of identification we use by showing three categories of schools. Our sample contains roughly 239,000 schools which are in PIN codes which never have ITES centers (or at least not ITES centers we observe). A further 172 schools are in PIN codes which have ITES centers, but do not add ITES centers during the survey period. Finally, we have 408 schools in PIN codes where the number of ITES centers changes over the course of the study. Given that our specifications will include school fixed effects, we are identifying off of these final 408 schools. In some specifications we will limit the comparison group to areas which are more comparable to those which have ITES centers.

In addition to this basic information on ITES center locations and founding dates, we undertook a follow-up survey of the centers in our sample. Although we attempted to survey all centers, in the end we were able to collect data on 83% (the remaining were missed largely due to refusal to answer survey questions). For these centers we have data on whether they operate in English, the number of employees and several employee characteristics. Information on number of employees and English-language operations is available for all the centers we surveyed; demographic information is available for a subset. The ITES centers are relatively small, with a median of 80 employees. Sixty-two percent of the ITES centers handle voice calls in English. Employees are young (median age of 28), largely without children and mostly from the local area.

As a final note, in addition to ITES centers within the same PIN code as the school, we use two variables measuring slightly further centers: those in PIN codes within 5km of the school's own PIN code and those 5-10km away. To calculate distance, we use GPS data on PIN code locations (the latitude and longitude are measured at the post office in each PIN code). We count the number of ITES centers in each of the two neighboring groups.

1.2.4 Placement of ITES Centers

A central issue in our analysis is the fact that ITES centers are not placed randomly. Our analysis will take advantage of variation over time, so any fixed differences across areas will be adjusted for, but it remains important to understand what drives placement.

We undertake two strategies. First, we can have an initial sense of the magnitude of this threat based on discussion with ITES center operators about location choices. The primary issues they cited when deciding where to locate were infrastructure and transportation: areas with no electricity and roads were not appealing places to operate. In addition, center operators cited their need to find high quality employees cheaply in the local area. There was some sense of a trade-off: there are more qualified individuals in larger cities, but people outside these areas demand lower wages. It is clear that center operators are thinking carefully about cost-benefit considerations. However, the central demographics discussed are very likely to be constant over time, at least over the short time frame of our study.

We are also able to evaluate this endogenous placement statistically using our data. To do this we estimate, at the neighborhood level: (a) the determinants of having an ITES center by the end of the sample in 2007 and (b) the determinants of adding an ITES center during the period we observe. We focus on variables cited by ITES center operators: whether the area has electricity, whether it is in a more urban area and whether there is an English-language school in the area. This last variable is intended to capture the availability of English-speaking individuals. We also include a control for total school enrollment and, in some cases, state fixed effects.

The results from these regressions are shown in Table 1.2. In general, the results support the interview evidence. More urban areas are more likely to have centers by 2007 and more

Table 1.2: *Placement of ITES Centers*

Dependent Variable:	*Number of ITES Centers, 2007*		*Add ITES Center During Sample*	
	(1)	(2)	(3)	(4)
Ever Had Electricity	-.0002	-.0001	-.0002	.0001
	(.0005)	(.0005)	(.0003)	(.0003)
Urban	.010***	.010***	.004***	.004***
	(.001)	(.001)	(.001)	(.001)
Any English School (0/1)	.007***	.006***	.003***	.003***
	(.0008)	(.0008)	(.0004)	(.0004)
Log Enroll. First Survey Year	-.00005	-.0001	.000001	-.0001
	(.0001)	(.0002)	(.0001)	(.0001)
State Fixed Effects	NO	YES	NO	YES
R-squared	0.002	0.003	.002	.002
Observations	71,667	71,667	71,890	71,890

Notes: This table shows the effects of PIN code characteristics on ITES center placement. The left hand side variable in Columns 1 and 2 is the number of ITES centers in 2007; in Columns 3 and 4 it is whether any centers were added during the sample period. Standard errors in parentheses, clustered at the PIN code level. *significant at 10% **significant at 5% ***significant at 1%.

likely to add them during the sample; these effects hold with and without state fixed effects. Areas with English-language schools are also more likely to have centers and more likely to add them during the sample; again, these results are robust to state fixed effects. We see limited evidence that electricity matters, although this may be due to the high correlation with urbanization; enrollment also does not seem to have any impact.

The inclusion of school fixed effects means that any differences in levels of enrollment associated with these variables will not impact our results. However, if there are differential trends in enrollment across neighborhoods associated with these variables, this could impact our results. To address this, in the results below we will allow for separate year fixed effects for areas that are more urbanized and areas with any English-language schools; this is discussed in more detail below.[12]

[12] We do not include separate trends in electricity or initial enrollment level since these do not impact placement; consistent with this lack of impact on placement, including these does not change our results.

1.3 Empirical Strategy

We estimate the impact of ITES centers on school enrollment using a fixed effects estimator. We observe enrollment in school i in PIN code j at year t; denote this variable n_{ijt}. In addition, we observe number of ITES centers in PIN code j at year t, which we denote c_{jt}. Our basic regression is shown in Equation (1) below

$$n_{ijt} = \alpha + \beta_1 c_{jt} + \gamma_i + \phi_t + \Psi X_{ijt} + \epsilon_{ijt} \tag{1.1}$$

where γ_i is a vector of school fixed effects and ϕ_t is a vector of date controls. These date controls include year fixed effects, and year fixed effects interacted with state fixed effects, neighborhood-level electricity, urbanization, school language and whether there is any English-language school in the area. Thus, we allow the year fixed effects to differ by state and by the variables that drive ITES center placement in Table 1.2. In addition to these fixed effects, we include a set of school-specific time-varying controls (X_{ijt}) measuring school-level infrastructure. The coefficient of interest is β_1, which captures the effect of ITES centers on school enrollment. This coefficient is identified off of schools in areas which add ITES centers during the sample. Throughout the analysis, we cluster our standard errors at the neighborhood level.[13] We will also estimate this overall regression including district-specific time trends.

Our left-hand side variable is the log of enrollment, allowing us to interpret our results as a percent change in enrollment. The use of the log form leads to the concern that our results could reflect movement between schools of different sizes. For example, if students leave large schools and move to smaller schools the log specification could show an increase even though total enrollment was stable. We address this by weighting our regressions by

[13] We choose to cluster at the neighborhood level (rather than at the school) since c_{jt} is the same for all schools within a neighborhood-year. In fact, the level of clustering makes relatively little difference – even clustering at the district level gives very similar standard errors. For example: in our primary regression, Column 1 of Table 3, the t-statistic with neighborhood clustering is 2.92 and with district clustering is 3.19, actually larger. In general, the significance of the results we show never change with district clustering. We should note that when we include district-specific trends in the regression we are *not* able to cluster at all given the large number of controls. This means the standard errors are likely biased downward in those regressions, although since the clustering does not make a large difference in general, it seems unlikely this bias is large.

number of students enrolled in the first year we observe the school, which gives greater weight to larger schools. In addition, in a robustness check we will run these regressions with the level of enrollment (count) on the left hand side.

A concern with estimating this equation on all areas is that our impacts might be identified off of rural areas which are not at risk of having ITES centers, and these may not be appropriate comparisons for those areas which get ITES centers. Given that, we will estimate, and focus on, a specification in which we limit the sample to areas which have at least one English-language school; these are the areas most "at-risk" for getting ITES centers.[14] As a further robustness check, we will also limit to areas which ever have ITES centers during the sample period. Although our primary results use fixed effects, in a robustness check we will show our central estimates in first differences.

As noted in the introduction, we are concerned about the possibility that the results are driven by other variables which are changing over time and influence both ITES centers and school enrollment. A related issue is the possibility that ITES center operators consciously introduce centers in places where school enrollment is increasing. To address both of these issues, we estimate whether *future* ITES centers predict current enrollment using Equation (2) below.

$$n_{ijt} = \alpha + \beta_1 c_{jt} + \beta_2 c_{j,t+1} + \gamma_i + \phi_t + \Psi X_{ijt} + \epsilon_{ijt} \tag{1.2}$$

$c_{j,t+1}$ is a variable measuring number of ITES centers in PIN code j in year $t+1$. If $\beta_2 > 0$ this would indicate that areas which get ITES centers next year have higher enrollment in this year, relative to their previous enrollment. This would point to ITES centers being introduced in areas which are growing faster. In contrast, a finding that $\beta_2 = 0$ indicates that ITES centers are not introduced into areas in which school enrollment is growing for other reasons. This technique has been used elsewhere to address this concern (Jensen and Oster, 2009; LaFerrara, Chong and Duryea, 2009). We also estimate Equation (2) including a trend for years until a new ITES center is introduced. This allows us to look slightly

[14] This is our strongest predictor of having an ITES center. Virtually all of our ITES centers are located in PIN codes which have at least one English-language school.

more generally at whether enrollments are increasing in years up to a new ITES center introduction. It is important to note that this technique does not allow us to rule out the possibility that ITES centers are introduced at exactly the same time as another innovation, and that the other innovation drives school enrollment. However, this possibility seems more remote.[15]

One important issue is the coverage of our ITES center dataset. Although we worked to cover as many ITES centers as possible, it seems extremely unlikely that coverage is perfect and there are likely areas that have ITES centers that we do not observe. This means that our "control" group of non-changers also contains some schools that should be in the "treatment" group. To the extent that there is a positive effect of ITES centers on school enrollment, this imperfect coverage should bias our estimates of β_1 downward, since the changes in the control group will be more biased upward by the inclusion of "treatment" schools.

1.4 Results: Impact of ITES Centers on School Enrollment

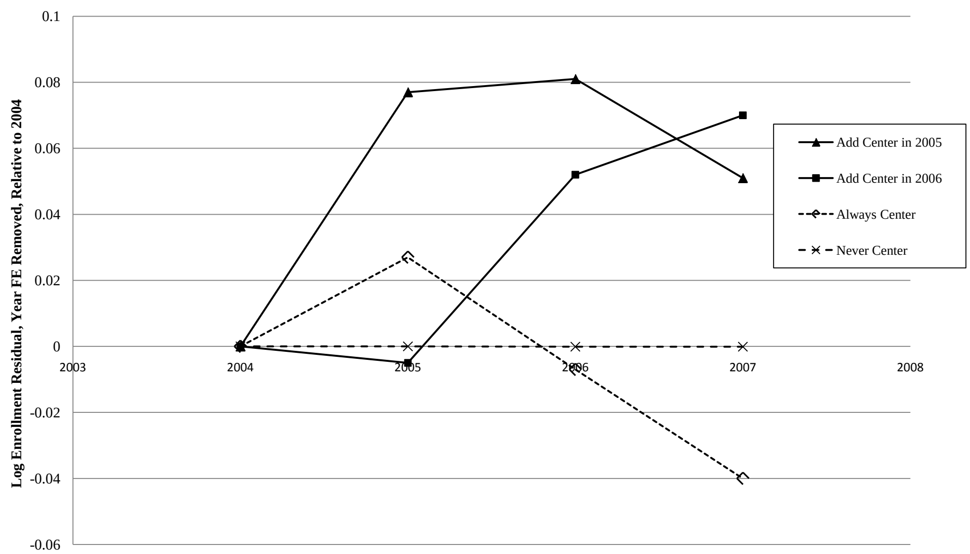
This section presents our estimates of the impact of ITES centers on enrollment.

1.4.1 Baseline Results

We begin by illustrating our results for a subset of ITES center introductions in Figure 2. To generate this figure, we focus on four groups of schools: (1) schools that always have an ITES center in their PIN code, (2) schools that add a center between the 2004-05 and 2005-06 school years, (3) schools that add a center between the 2005-06 and 2006-07 school years and (4) schools that never have any ITES centers. Note that this is not the universe of schools which add ITES centers during this period, but it includes many of the changers. We focus

[15]We also cannot rule out the concern that ITES center operators are targeting areas which seems like they would have large enrollment *responses* to these centers. Under this theory, our results would be valid within sample but would overstate out-of-sample effects. This would require, however, that ITES center operators are choosing locations based on the elasticity of primary school with respect to future returns. This seems unlikely given the difficulty of measuring these parameters and the insignificant effect they would have on short- and medium-run outcomes for the firms. Our conversations with ITES center operators also gave no indication of this type of consideration in placement.

Figure 2:
Impact of ITES Centers on School Enrollment



Notes: This figure shows changes in enrollment over time for four balanced panels of schools. All enrollment numbers are residuals from a regression of log enrollment on year fixed effects and values represent changes relative to the residual values in 2004. The year 2004 refers to the 2004-2005 school year (beginning in June 2004, ending in April 2005); enrollment is recorded as of September 2004. Schools are coded as adding a center in 2005 if an ITES center is founded in the area any time during 2005.

Figure 1.2: *Impact of ITES Centers on School Enrollment*

on this group because we can isolate a balanced panel of schools which are observed for four years (2004 through 2007). Using this sample of schools, we regress log enrollment on year fixed effects and take the residuals; this removes any consistent year-by-year variation. These residuals are graphed in Figure 2, which shows changes in these residuals relative to the level in 2004.

The key result in Figure 2 is that there are large year-on-year changes in enrollment in the two groups that add ITES centers during the sample, and these changes line up in terms of timing with the ITES center addition. In areas that add a center between 2004 and 2005, schools see a large increase in enrollment between these years, whereas there is only a small increase in schools that always have centers, and no change for schools that add centers later or never add them.[16] Further, for areas that add an ITES center between 2005 and 2006 there is a large increase in enrollment between these years, but no change in the year before. This is the only group with a large increase between 2005 and 2006. Overall, the figure demonstrates large changes in enrollment which correspond to ITES center introductions.[17]

In the regressions, we return to our whole panel, which includes areas which add centers in years other than 2005 and 2006. Panel A of Table 1.3 shows our statistical estimates of the effect of ITES centers on enrollment. Column 1 presents our results using the entire sample. The coefficient on ITES centers is positive and significant: adding one more ITES center increases school enrollment by 4.6%. Column 2 shows this regression with district-specific trends included, to address the concern that districts that have ITES centers introduced are trending differently than those that do not. The coefficient is slightly smaller (3.0%) but still highly significant. Our preferred estimate appears in Column 3, in which we limit to areas with at least one English-language school, which means our non-changer areas are most comparable to the areas which add ITES centers. Although this restricts the sample significantly, the coefficient is similar in magnitude to the overall sample (4.8%) and highly significant. Finally, Column 4 limits further to areas that ever have an ITES center that we observe (including those that change and those that always have a center). The coefficient is again even larger and significant, despite the extreme sample size restriction.

In Panel B of Table 1.3 we explore whether the introduction of ITES centers in the slightly broader surroundings matter. As described, we do this by estimating the impact of ITES centers in neighboring PIN codes. We focus on those in PIN codes very close by (within 5 kilometers) and those slightly further (5-10 kilometers away).[18] Panel B demonstrates that there are some impacts for ITES centers in the nearest neighbors. Focusing on our preferred

[16]As shown in Table 1.1, most schools never have any ITES centers. For this reason, the year fixed effects are largely identified off of these areas, so when we generate residuals removing these fixed effects, the average residuals in these area are very close to zero.

[17]It is not clear why those schools which always have a center have an increase and then a decease in enrollment; this may reflect different yearly conditions in these areas versus those who never have a center.

[18]Distance is measured from center-to-center of the PIN codes, so it is possible that a given individual may be closer or further away, but should be in a similar range.

Table 1.3: *Effect of ITES Centers on School Enrollment*

Dependent Variable:	*Log Enrollment*			
Panel A: Number of ITES Centers in PIN Code				
Sample:	All Schools		In PIN Code with Any English Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
	(1)	(2)	(3)	(4)
Number of ITES Centers	.046***	.030***	.048***	.071***
	(.016)	(.010)	(.015)	(.014)
Observations	911,499	911,499	314,476	2,121
Panel B: Number of ITES Centers in Neighboring PIN Codes				
Sample:	All Schools		In PIN Code with Any English Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
	(1)	(2)	(3)	(4)
ITES Centers in PIN Code	.038***	.024**	.040***	.075***
	(.011)	(.010)	(.012)	(.018)
# ITES Centers in PIN Codes	.002**	.002***	.002**	.002**
within 5km	(.0007)	(.0005)	(.001)	(.0006)
# ITES Centers in PIN Codes	.0002	-.0002	.0002	.00003
5-10km away	(.0004)	(.0002)	(.0004)	(.001)
Observations	911,499	911,499	314,476	2,963
Panel C: Impact of Future ITES Centers				
Sample:	All Schools		In PIN Code with Any English Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
	(1)	(2)	(3)	(4)
ITES Centers	.046***	.033**	.050***	.072***
	(.016)	(.014)	(.017)	(.019)
ITES Centers Next Year	-.001	-.004	-.003	-.001
	(.013)	(.012)	(.014)	(.016)
p-value, This Year=Next	.06	.11	.05	.02
Observations	911,499	911,499	314,476	2,121

Standard controls: School fixed effects, time-varying school plant characteristics, year dummies interacted with dummies for state, urban, school language, and English language school in PIN code. District Trend controls: Standard controls plus district-specific trends.

Notes: This table shows our primary estimates of the impact of ITES centers on school enrollment. The independent variable measures the number of ITES centers in the same PIN code as the school. Columns 1-2 include all schools. Column 3 is limited to PIN codes with any English schools. Column 4 is limited to schools which ever have an ITES center in their PIN code (either always have the same number or change during the sample). Standard errors (in parentheses) are clustered at the neighborhood level in Columns 1, 3 and 4; clustered errors could not be estimated when district trends are included in Column 2. *significant at 10% **significant at 5% ***significant at 1%. All regressions are weighted by initial school enrollment level.

specification in Column 3, we find one more ITES center in one of the closest neighboring PIN codes has a significant but very small impact (around 0.2%). ITES centers in more distant PIN codes (5-10 km away) have no impact. This suggests effects are extremely localized.

The evidence in Table 1.3 suggests a strong connection between ITES centers and total number of children in school. In Online Appendix Table 1 we show these effects broken down by demographic group. We find the effects are similar for girls and boys.[19] We also see similar effects in the three states in the sample, although the effect is significant only in Karnataka and Tamil Nadu. Perhaps most informative, the impacts are largest for older grades. This suggests that much of the impact may be due to children staying in school rather than newly enrolling at the youngest ages.

In Online Appendix Table 2 we show a number of additional robustness checks. The first panel shows the results in first differences rather than fixed effects. The second shows the effects from Columns 3 and 4 with district trends included. The third panel shows impacts on enrollment levels rather than changes. The fourth panel estimates the impact of having any ITES center, rather than the number of centers.

The results look very similar in all of these robustness checks. An exception is the analysis, in Panel D, of the impact of any ITES center. Although these impacts are positive they are generally not significant. Practically, this is likely due to the fact that we have much less scope for identification in these cases. But, broadly, this suggests that adding more centers matters even if there is already a center around, which may reflect the extreme localization of information – even if there is already a center in your PIN code, getting one closer to you may matter. We will explore this in more detail in Section 6.

In a final analysis, the last panel of this appendix table limits to *only* PIN codes which add a center in 2005 or add a center in 2006, which means the regressions are identified off of the differential timing of changes. In this – essentially our most stringent specification –

[19]The fact that the impact for boys and girls is similar may seem puzzling, given the focus on the female nature of this work. In fact, in the ITES centers in our data, slightly less than half of the employees are women, which may explain the similar impact.

we still see a large and significant impact of ITES centers on enrollment.

Future ITES Centers

The central threat to the validity of our estimates is the possibility that ITES center introduction anticipates schooling increases rather than causing them. This is related to the issue of endogenous ITES center placement. As discussed above, to the extent that endogenous placement reflects only characteristics which are constant over time this will not drive our results since we include school fixed effects. Further, if trends are different for areas which are urban, or have more English-language schools, we have also addressed this issue. The concern which remains unaddressed in our main specification is the possibility that ITES centers are located in areas that are changing in other ways that we do not observe. There are at least two specific concerns. One is that ITES centers are placed in areas where schooling is increasing more quickly, since center operators are targeting a future labor force (given that our estimates are for primary schools, this would be a fairly distant future). A second concern is that some other unobserved variable ("modernity", for example) drives both ITES center introduction and school enrollment.

To address this concern directly we estimate whether future ITES center placement predicts current enrollment. If it does, this would suggest ITES center introduction anticipates changes in schooling, rather than causes them. Panel C of Table 1.3 replicates Panel A, but includes a control for the number of ITES centers in the following year in addition to the indicator for current ITES centers. Adding the control for future ITES centers has only a small impact on our estimates of the effect of current ITES centers. In addition, and more importantly, the effect of future ITES centers is small and not statistically precise, suggesting no strong evidence of pre-trends. In general, we can reject equality between the coefficients on current and future ITES centers.

In Online Appendix Table 3 we do a similar test, but rather than simply controlling for having an ITES center next year, we control for a time trend up to the year of ITES center introduction (the trend is defined so higher values indicate the center introduction is closer in time). If ITES centers are introduced into places where enrollment is increasing more quickly, we should see evidence of a positive trend. We do not see this. The trend coefficients are small and not significant. It is important to note that the results here do not indicate that ITES center placement is exogenous, but instead indicate that this endogenous placement does not drive our results.

1.4.2 Impacts of ITES Centers by Language of Instruction

The evidence above suggests that overall school enrollment increases in response to ITES center introduction. Here, we turn to separating the result by language of instruction. One of the central features of ITES centers in India is that the vast majority operate in English. In our survey, all of the voice ITES centers (which make up about half of our sample) use English; the majority of non-voice centers also require English. Given this, to the extent that what we observe reflects changes in schooling in response to job opportunities, these changes should disproportionally result in higher English-language school enrollment.

We separate our effects by language of instruction in Table 1.4. We generate new variables interacting the number of ITES centers with language of instruction and control separately for the impact of ITES centers on local language schools and on English-language schools.[20] Panel A of Table 1.4 shows our basic test of differences across school types. Column 1 reports impacts on total enrollment using the entire sample. We find the total impact of ITES centers in English-language schools is large and significant; the impact of ITES centers in local-language schools is smaller and not significant. In our preferred specification (Column 3), we find that enrollment in English-language schools increases by 7.1% for each ITES center introduced. The p-values reported at the bottom of the table indicate we can reject the equality of the impacts in the two school types. One thing which is important to note is that we *do not* see decreases in enrollment in local language schools. The increase in enrollment in English-language schools does not appear to come at the

[20] The two variables are mutually exclusive; each coefficient can be interpreted as the effect for that school type.

expense of enrollment in local-language schools.

In Panel B of Table 1.4 we push the data on language further, and separate schools into three groups: those that do not teach in English at all, those that teach some in English and some in another local language and those that teach only in English. Consistent with the larger impact for English-language schools overall, we find the effects are largest for schools that teach exclusively in English. However, the difference between these and those that teach partially in English are small and inconsistently signed. The largest distinction appears to be between schools that teach at least some English and those that teach none.

Finally, in Panel C of Table 1.4 we test for impacts of future ITES centers separately by language of instruction. Future ITES centers do not impact enrollment in either local-language or English-language schools. The coefficients on future centers are small and inconsistently signed. We should note that we generally cannot reject equality between the coefficients (other than in Column 4); this is due to the more limited precision on the estimates when we separate the schools into the two groups. The p-values do approach marginal significance.

1.5 Robustness: Changes in Population and Income

This section addresses several key robustness issues. In particular, we evaluate whether it is possible that our results are simply driven by mechanical changes in number of schools, population or income deriving from the ITES center introduction.

1.5.1 Changes in Population

A key downside of our data on education is that we observe number of students enrolled, not enrollment rates. This introduces the possibility our results could be driven by population increases. The controls thus far address the concern that ITES centers are introduced to more populous areas or areas which are growing faster. However, if the ITES center itself increases population, this could produce our result. This would be a concern if we were, for example, considering the impact of introducing a large manufacturing plant to an isolated

Table 1.4: *Effects by Language of Instruction*

Dependent Variable:	*Log Enrollment*			
Sample:	All Schools		In PIN Code with Any English Schools	Ever Had an ITES Center
Panel A: Impact of ITES Centers by School Language of Instruction				
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
# Centers × Local Lang.	.019	.004	.020	.044**
	(.019)	(.015)	(.021)	(.018)
# Centers × English	.073***	.057***	.071***	.102***
	(.016)	(.015)	(.017)	(.020)
p-value, English=Local Language	0.01	0.01	0.03	0.02
Observations	911,499	911,499	314,476	2,121
Panel B: Impact of ITES Centers by Detailed School Language of Instruction				
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
# Centers × Local Lang.	.019	.004	.020	.045**
	(.019)	(.015)	(.021)	(.018)
# Centers × Some English	.072***	.061***	.069***	.096***
	(.0169)	(.015)	(.018)	(.024)
# Centers × All English	.074***	.051**	.073***	.118***
	(.030)	(.023)	(.032)	(.033)
Observations	911,499	911,499	314,476	2,121
Panel C: Pretrends by Language				
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
# Centers × Local Lang.	.031	.018	.036	.053**
	(.021)	(.019)	(.024)	(.019)
# Centers × English	.065***	.052***	.065***	.097***
	(.023)	(.020)	(.023)	(.032)
# Centers Next Year	-.018	-.019	-.022	-.017
× Local Lang.	(.017)	(.018)	(.020)	(.016)
# Centers Next Year	.009	.005	.007	.007
× English.	(.023)	(.016)	(.023)	(.022)
p-value this Year vs. Next, Local	0.11	0.26	0.11	0.02
p-value this Year vs. Next, English	0.20	0.16	0.18	0.09
Observations	911,499	911,499	314,476	2,121

Notes: This table shows the impact of ITES centers by school language of instruction. Panel A shows the differential effects for English and local language schools. Panel B shows the effect for local language schools, schools with some English instruction, and schools with exclusive English instruction. Panel C shows the effects by school language for voice and non-voice ITES centers. Voice centers are defined as ITES centers where at least half of employees handle voice calls. Columns 1-2 include all schools. Column 3 is limited to PIN codes with any English schools. Column 4 is limited to schools which ever have an ITES center in their PIN code. Standard errors (in parentheses) are clustered at the PIN code level in Columns 1, 3 and 4; clustered errors could not be estimated when district trends are included in Column 2. For standard controls see Table 1.3 *significant at 10% **significant at 5% ***significant at 1%. All regressions are weighted by initial school enrollment level.

area. In this case, however, we argue this is unlikely to explain more than a very small fraction of the effect we observe.

To begin, it seems appropriate to calibrate the magnitude of our results in terms of the change in number of students. Focusing on our preferred specification in Column 3 of Table 1.3 we find a 4.8% increase in enrollment after the introduction of an ITES center. Based on a median school size of 143, this is 7 students per school, which aggregates to about 180 students in the PIN code overall.

The first question is whether in-migration among the employees of the ITES center themselves could be driving this change. There are several reasons we think this is unlikely. First, ITES centers tend to employ young, childless individuals. This can be seen in anthropological and ethnographic work on ITES centers in India (i.e. Ng and Mitter, 2005) and directly in our ITES center survey data. In the average center in our sample, managers reported that 10% of employees have children, so the potential increase in children in the area even if all employees were new to the area is small. Second, relocation for work in ITES centers is also relatively rare (12.2% of employees). Even if we assume *all* this relocation is by people with children we find an average of 5.6% employees with children relocate. At the median ITES center, which has 80 employees, this amounts to just 4.4 people with relocated children. This can be compared to the 180 student increase we estimate for the PIN code overall with a single ITES center introduction. In fact, this number is likely to be an upper bound; in reality, the individuals with children are generally the *least* likely to relocate.

There remains a concern that the introduction of an ITES center may bring with it other service jobs, which could increase population.[21] This could mean other jobs in the ITES center itself (although this should be captured in our employment measure) or, more likely, jobs working for ITES center employees (e.g. drivers, maids, cooks). If people migrate into the towns for these jobs, this could result in population changes. The first argument

[21] It is also possible that the introduction of an ITES center is associated with an overall increase in other types of businesses, which bring in more migrants. The evidence in Panel C of Table 1.3 above limits this concern; for this to drive our results, it must be the case that these other businesses enter at exactly the same time. This is also largely addressed by our evidence following this paragraph.

against this is again calibration-based. As we note, the total student increase is about 180, and the median ITES center employs 80 people, of whom about 12% migrate in. If we assume that relocation for work in support jobs is as frequent as for work in the ITES centers themselves[22], then each ITES center employee would have to bring with them, directly or indirectly, 18 children between the ages of 6 and 14. Put differently, if we assume that each ITES center worker hires two new servants, and 10% of those individuals migrate in with two school-aged children, this would explain about 15% of our effect, still very small despite the fact that these are very generous assumptions.

As a second, related, calibration, we note that migration in general in these states in India is fairly limited. In other household survey data (the 1998 and 2005 National Family and Health Survey (NFHS)) we can estimate what share of school-age children report migrating to a new area within the last year. For 84% of clusters in the NFHS sample there are zero in-migrants in the last year among school-aged children.[23] Even the clusters at the 90th percentile on this measure still have only 5% of school-aged children who are in-migrants in the last year.

In addition, we show two analyses which test the population mechanism directly. First, in Online Appendix Table 4 we show, for the subset of areas for which the school reports total neighborhood population, the impact of controlling for population on our results (these population data are recorded by the school, and vary across years).[24] We do not want to lean very heavily on the evidence in these regressions since we observe population only for a small subset of the sample and it is unclear how the school estimated population. However, this table demonstrates that including a control for population in the regressions does not significantly impact our estimates. The coefficients are noisier, but this seems to be due to

[22]In fact, this is an overstatement: based on nationally representative data from India, migration is least likely among low-skill individuals and most likely among those with high skills.

[23]In the NFHS a survey cluster typically covers a single village or area within a town and includes a randomly selected subsample of individuals.

[24]Since there are multiple schools in each area, we cannot generate enrollment rates off of these data, since the population reported is an area-level population not simply the population relevant for that school. The fact that this is true should also be clear from the coefficient on population; it is much lower than one, which at least partially reflects the fact that as the area population increases, not all of that increase goes to a given school.

Table 1.5: *Robustness: Number of Employees*

Dependent Variable:	*Log Enrollment*			
Sample:	All Schools		In PIN Code with Any English Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
	(1)	(2)	(3)	(4)
Number of ITES Centers	.073***	.062***	.069***	.099***
	(.017)	(.016)	(.017)	(.018)
Log Number of Employees	-.0002	-.0002***	-.0001	-.0002**
	(.0001)	(.0001)	(.0001)	(.0001)
Observations	911,499	911,499	314,476	2,121

Standard controls: School fixed effects, time-varying school plant characteristics, year dummies interacted with dummies for state, urban, school language, and English language school in PIN code.
District Trend Controls: Standard controls plus district-specific trends.

Notes: This table shows the effect of the number of employees of a call center on school enrollment. The independent variables measure the number of ITES centers in the same PIN code as the school, and the natural log of the number of employees in the center, respectively. Columns 1-2 include all schools. Column 3 is limited to PIN codes with any English schools. Column 4 is limited to schools which ever have an ITES center in their PIN code (either always have the same number or change during the sample). Standard errors (in parentheses) are clustered at the PIN code level in Columns 1, 3 and 4; clustered errors could not be estimated when district trends are included in Column 2. *significant at 10% **significant at 5% ***significant at 1%.

changes in the sample: there is very little difference between Panel A (restricted sample only) and Panel B (restricted sample, with population control).

Second, we note that an important implication of any mechanical mechanism like migration is that larger ITES centers should have larger effects on enrollment. This need not be the case if the effects are driven by changes in information provided by the existence of the ITES center. In fact, we find that the impacts we observe **do not** scale with the size of the ITES center. In Table 1.5 we show our primary analyses but include in the regression a control for number of employees.[25] If enrollment effects were scaling with the size of the ITES center, the coefficient on employees should be positive, but it is negative and indistinguishable from zero in all specifications.

As a final argument, we note that if migration was driving the increase in school

[25]We collected this information for most, but not all, of our ITES centers as detailed in Section 2.

enrollment, we would expect to see similar gains across age groups in schooling, since it would just reflect the ages of the workers' children. As we discussed previously, increases are not homogeneous across age groups, but in fact are larger for older students, which is inconsistent with the population-driven story.

ITES Center Driven Changes in Number of Schools A related, but more minor, concern is that our results are driven by changes in the number of schools in the area. If the introduction of an ITES center causes a decrease in the number of schools then the remaining schools could see enrollment increases even if the total enrollment rate in the area remains constant. We evaluate this by estimating the impact of ITES center introduction on the count of schools in the neighborhood. Estimates are shown in Online Appendix Table 5. The results indicate that changes in school count are not a concern: the impact on number of schools is very small and not significant. We also demonstrate that the introduction of ITES centers does not affect the number of English-language schools or school infrastructure.

1.5.2 Changes in Income

A second concern with our results is the possibility that ITES centers drive enrollment because they increase income and schooling is a normal good. We note that this seems unlikely, given the results in Table 1.5, which show no impact of number of employees on enrollment effects, since the total income increase should be greater for larger ITES centers. Still, we take advantage of the fact that existing literature has provided estimates of the income elasticity of school enrollment in similar contexts to estimate the magnitude of predicted enrollment increase resulting from increased income from ITES centers.

A number of papers have estimated the income elasticity of school enrollment in the developing world (Alderman et al., 2001; Glick and Sahn, 2000; Glewwe and Jacoby, 2004; Orazem and King, 2007); these estimates range from 0.25 to 1.25. In Online Appendix A we go through a detailed calculation of the impact of ITES centers on area level income; we estimate an effect of around 0.57%. Combining these figures and comparing to the overall impacts we estimate on school enrollment, the income changes could account for only a small share of the changes in enrollment. Even at the largest elasticity estimate, this figure is only around 15%.

Similar to the case of population, an auxiliary concern is that the ITES center brings other businesses, which also increase income. It is more difficult to rule this out than in the population case. However, the fact that we do not see evidence of pre-trends suggests that these new businesses would need to arrive at exactly the same time as the ITES centers. In addition, given the very small share of the effect which is plausibly explained by ITES center income, in order for income overall to explain a larger share, these other businesses would need to swamp the ITES centers in their income contribution, which seems unlikely.

It is important to note that our argument in this section is *not* that migration and income changes play no role in our results but simply to say that calibration suggests these effects are small.

1.6 Mechanisms: Localized Information versus Localized Returns

We draw several conclusions based on the results in Sections 4 and 5. The introduction of an ITES center to an area results in an increase in school enrollment and this increase is concentrated in English-language schools. The observed increase does not appear to be driven by mechanical changes in the number of schools, population or income. Finally, these changes are very localized: ITES centers even slightly further away have little or no impact on enrollment. Based on these results, we argue that the effects we observe reflect responses to changes in the perceived returns to schooling after the introduction of new local job opportunities.

In this section we provide some initial evidence on the mechanisms that drive this effect. We distinguish two possibilities. First, the introduction of an ITES center may impact actual returns to schooling by providing new jobs at *that* center. Alternatively, it may impact perceived returns to schooling by providing better information about these jobs in general, even if the change in actual job opportunities is limited. This distinction is potentially important for thinking about policy implications. In this section we use a supplementary dataset which we collected in Madurai District (in Tamil Nadu) to provide some evidence on this question.

To fix ideas, consider the simplest model of schooling decision-making in a context with no information frictions. Assume there are two locations, A and B, both of which begin with no ITES centers and otherwise identical job opportunities and education costs. Assume education is a binary choice which carries some positive wage returns. At some date, an ITES center is introduced into area A and (because we are assuming information is shared fully) it is immediately observable to individuals in both areas. The existence of this center increases the wage returns to education while education costs remain the same.

For individuals in area A, the value of education increases by the full amount of the increased wage returns. For individuals in area B, however, the increase is less because to take advantage of the new jobs, they would need to migrate to area A. Assuming the cost of migration is positive, the reaction of individuals in area B to the ITES center should be smaller than in area A; how much smaller depends on migration costs. Note that these migration costs could be the cost of moving to live in a new area, or the cost of travel to work in that area.

Now consider adding information frictions so the information about the increased returns diffuses only partially (or not at all) between areas A and B. In this case, the response in area B will be less than in area A *even if costs of migration are small*; how much less will depend on how limited information diffusion is. This suggests two peices of information are key to distinguishing these models: the extent of migration (for work and for children leaving home) and the localization of information.

1.6.1 Survey Data from Madurai

We fielded a survey in Madurai District in Tamil Nadu. Madurai is a small city about 450 kilometers from Chennai with several ITES centers. We surveyed 1000 individuals: 500 in Madurai itself and 250 in each of two smaller towns, Thirumangalam and Peraiyur, 20 to 50 km away. We will focus on the Madurai data for this analysis. We collected data including distance to work, future plans for children and knowledge of ITES centers. Importantly, we collected GPS data on location of households and ITES centers, allowing us to calculate exact distances. Details of the survey appear in Online Appendix B.

Evidence on Costs of Travel and Migration

The Madurai data do not support the claim of very limited migration. The median person in our sample who is working reports working 2 kilometers away from where he lives; 25% work more than 6 kilometers away. Among people with at least ten years of schooling – presumably most likely to work at high wage jobs like those at ITES centers – the median person reports working 4 kilometers away and 25% work more than 10 kilometers away. This suggests that it is not unusual to travel reasonable distances to work.[26] Children migrating away from home as adults is even more common. Among children of the sample participants who are over 18, roughly 40% of them live away from home, and 25% live more than 5 kilometers away.

The evidence on migration is echoed by larger datasets. Data from the National Family and Health Survey show that among working individuals ages 20-35 with at least a secondary school education, roughly 30% have moved in the last five years. Similarly, in the 2001 Census, 29.9% of all persons were living in a town other than that of their birth.

Taking this evidence together, it seems unlikely that our DISE results reflect localized changes in actual returns.

Evidence on Information Diffusion

We turn now to patterns of information diffusion. We focus on relating distance to an ITES center (calculated based on GPS coordinates) to two pieces of data reported by the households: knowledge about ITES centers and whether parents plan on ITES center jobs for their children.

[26] As a side note, this also supports the argument in Section 5 that our estimates are not driven by population or income. Since people travel for work, any income impacts would be less localized than the ITES center impacts we estimate.

Knowledge of ITES Centers We focus on five knowledge variables: whether the respondent reports knowing anyone who works in an ITES center, whether they report that there is an ITES center within the "local area"[27] and three measures of their knowledge about ITES center job qualifications. The job qualification questions listed a set of characteristics (e.g. speak English, college graduate) and asked individuals whether these were "required" for jobs in an ITES center; in some cases the correct answer was yes, and in others it was no. We generate three measures of knowledge: the share of questions for which individuals reported they "didn't know" whether the qualification was required, the share of the true qualifications they correctly identified and the share of the false qualifications they correctly identified.[28] Online Appendix B reports summary statistics on the variables.

We begin by looking at how information varies by distance category. Panel A of Table 1.6 estimates coefficients on two dummies: being within a half a kilometer of the closest ITES center and being between 0.5 and 1.5 kilometers away; the omitted category is between 1.5 and 3 kilometers away. We see that knowledge is the highest in areas within a half a kilometer of the ITES center on four of the five measures, which is consistent with the evidence from the DISE data of effects decaying over relatively short distances.[29] The exception is when we explore impacts on the share of people who correctly identify true qualifications where nearly everyone gets a perfect score.

Online Appendix Figures 1-5 show smoothed plots of the knowledge outcomes against distance from the closest ITES center. There is strong evidence that information deteriorates quickly in the area right around the ITES center. Between 0 and 2 kilometers, moving

[27]The definition of "local" was up to the respondent.

[28]The correct qualifications were: speak English, use a computer and be a college graduate. The incorrect qualifications were: politically active, have a driver's license and be a woman.

[29]Though the negative effect for the middle category (0.5-1.5 km away) is somewhat curious, it is likely being driven by an outlying center. Honeywell, by far the largest ITES center in Madurai, is located on the outskirts of town. Unlike the smaller centers, people close to Honeywell all have relatively high basic knowledge of what ITES centers are, and have heard of someone who works there. The deeper knowledge, however, such as qualifications for working there, decays at the same rate as for other ITES centers. This is likely due to the fact that Honeywell is simply much more visible in the neighborhood than a typical ITES center in our sample. Because Honeywell is located relatively far outside the city center, everyone close to it is more than 1.5 km away from any other ITES center. Thus, a large percent of the people who live near enough to Honeywell to be impacted by it actually fall into the "farthest" category in these regressions.

Table 1.6: *Knowledge of ITES*

Panel A: Knowledge of ITES by Distance to Closest, Within Madurai					
	Heard of Someone who works at ITES	Know of ITES in local area	% ITES Ques. Answer "Don't Know"	% True Qual. Answer Correct	% False Qual. Anwert Correct
ITES <0.5 km	.212***	.230***	-.099**	-.015	.265***
	(.082)	(.078)	(.042)	(.042)	(.061)
ITES 0.5-1.5 km	-.129***	-.110***	-.009	-.021	.015
	(.040)	(.037)	(.041)	(.021)	(.030)
Controls	YES	YES	YES	YES	YES
Observations	494	496	498	498	498
Panel B: Knowledge of ITES by Distance to Closest, Within 1 km of an ITES Center					
	Heard of Someone who works at ITES	Know of ITES in local area	% ITES Ques. Answer "Don't Know"	% True Qual. Answer Correct	% False Qual. Anwert Correct
Distance to ITES	-.641***	-.488***	.215***	-.052	-.445***
	(.130)	(.131)	(.065)	(.070)	(.111)
Controls	YES	YES	YES	YES	YES
Observations	137	139	139	139	139
Panel C: ITES as Job Possibility for Child					
Dependent Variable: ITES listed first as possible job for child					
	All Children	Ages <10	All Children	Ages <10	
<0.5 km to ITES	-.004	.167**	-.072	-.035	
	(.065)	(.080)	(.075)	(.095)	
<1 km to ITES	-.0008	.026	-.063	-.024	
	(.039)	(.042)	(.049)	(.051)	
In English School			-.100*	-.092	
			(.057)	(.060)	
<0.5 km ×Eng.			.292*	.586***	
			(.153)	(.168)	
<1 km ×Eng.			.175**	.129	
			(.081)	(.082)	
Observations	170	81	169	81	

Controls: Child age, child sex, head of household education, whether respondent reports call center as highest wage job, asset ownership(television, radio, refrigerator, toilet).

Notes: Data comes from the survey run in Madurai District, Tamil Nadu. All regressions are limited to households within Madurai. In Panels A and C, the omitted distance category is more than 1.5 kilometers away. Dependent variables are the same in Panels A and B (abbreviated in Panel B). Columns 3-5 rely on answers to a set of six questions about qualifications which are "required" for job in an ITES center. Details are in Section 6. Controls: head of household education, whether the household head speaks any English and number of assets held by household (television, radio, refrigerator, and toilet). Standard errors in parentheses. *significant at 10% **significant at 5% ***significant at 1%.

further away decreases knowledge. Consistent with the estimate on the second dummy in the regressions in Panel B, there is some evidence that people who live much further away (between 2 and 3 kilometers) have better information.

In Panel B of Table 1.6 we take this analysis one step further and estimate the impact of distance to an ITES center on knowledge within 1 kilometer. As we squeeze the data in on a smaller area we increase the comparability across individuals, as well as the comparability across the ITES centers to which they are exposed. Despite the smaller sample size, we see a highly significant relationship between distance and knowledge. Individuals who are closer to an ITES center (so distance is smaller) are more likely to report knowing someone who works at one of these businesses, and more likely to report one in the local area. Further, those who are closer to an ITES center are less likely to report they don't know what qualifications are required and more likely to reject the false qualifications.

Child Job Choices Our second piece of evidence on information focuses on job choices for children. In the survey, we asked individuals about the most likely jobs for their child; they were given a list of possible jobs and asked to list three options. We focus on whether they choose the job "Call Center/BPO Worker" as the most likely job and analyze how proximity to an ITES center impacts this outcome. Since enrollment declines as children age, there is more selection in the older sample; given this, we run regressions on the whole sample and limited to children ages 5-10.

Panel C of Table 1.6 reports regression results from the Madurai-only sample. Column 1 uses the entire sample and estimates coefficients on the two distance dummies; controls are child sex and age, head of household education and whether the respondent reports that "Call Center/BPO worker" is one of the three listed jobs with the highest wages. This first regression shows no evidence that proximity to ITES centers matters for whether parents envision this job for their children. In Column 2, however, when we limit to younger children we see a strongly positive impact of being close to an ITES center.

The difference across age groups could reflect differential selection. It is also possible that this difference reflects the fact that schooling choices are more malleable for younger children – for example, it might still be possible to switch them to an English school. To explore this, in Columns 3 and 4 we interact distance with whether the child is enrolled in an English-language school (controlling for the overall English-language impact). These results are more striking. For both the overall sample and for the younger children, we observe that for children enrolled in an English-language school, proximity to an ITES center strongly impacts whether the parent reports that an ITES center job is likely. The fact that this occurs for the overall sample in addition to the younger children suggests that the lack of impact for the total sample in Column 1 is due to lack of language flexibility among older children.

Overall, this table suggests that there is an increase in perceived chance of ITES center jobs for children when an ITES center is closer. Again, this points to a very sharp decay of information about these jobs even over small distances.

Returns to Schooling and Reported Changes in Behavior As a final note, we present two more speculative pieces of evidence that are supportive of the information story. The first is on returns to schooling. We asked individuals their "best guess" about the monthly wage in the area for someone with a secondary school degree and for someone with only primary school; we define "returns to schooling" as the simple difference between these two values. Areas within half a kilometer of an ITES center report monthly returns 350 Rs higher than more distant areas and this effect is significant (tables available from the authors).

The second piece of evidence comes from the last question on the survey. For the 131 individuals in the sample who reported knowing of an ITES center in the local area, we asked whether they had made any change in response to that center introduction. Of course, it is extremely difficult to interpret responses to questions like this, especially given that it was asked at the end of the survey, which leaves open concerns about priming. However, the results are striking. About 50% report intentions to increase schooling for their children, some of whom cite specifically that they will enroll their children in English-language schools. It is interesting to note that this is the only behavior change reported – there is no mention of individuals themselves getting jobs at ITES centers – which is consistent with the evidence in Section 5 that these centers probably do not have large impacts on current income.[30]

We argue that the evidence in this section suggests that the localized impact of ITES centers that we observe in the DISE data reflect limited information diffusion rather than localized labor markets. It is worth noting that, in addition to being complementary to our findings above, this evidence is also quite complementary with experimental evidence on the role of information in schooling – in particular, Jensen (2010) and Jensen (2012). Both of those papers suggest that experimentally varying information about returns to schooling impacts schooling choices, and that is true even though in principle the information is available. The evidence here indicates that very limited information diffusion in the non-intervention context may explain why a simple provision of information can be so powerful.

1.7 Conclusion

In this paper we argue that the introduction of ITES centers in India has large impacts on school enrollment, and these impacts are concentrated in the very local areas around the ITES centers. We argue this effect is causal, and is not driven by pre-trends or mechanical changes in area-level population or income. The very local nature of our analysis and the fine timing of the effects are helpful in ruling out the concern that endogenous placement or trends in unobservables drive the impacts we see. Further, we provide some suggestive evidence that the very localized nature of the impacts may reflect limited information about non-local job opportunities; we argue this is more likely than the claim that these new job opportunities only impact local returns to schooling.

[30]Of 131 individuals who know of an ITES center in their local area, when asked if they will change their behavior because of the ITES center, 47 answered no, 9 answered that they will make their child study more or longer, 19 answered that they will make their child learn English, 38 answered that they will make their child learn computer skills or typing, and 18 indicated that there will be a change, but do not specify further. Not one parent answered that they would try to get a job at this center, or a nearby business, or any other change that was not related to investment in human capital for their child. While there is some danger of priming with this question (it was asked at the end of a survey about education and ITES centers, among other things), it is consistent with our assertion that these ITES centers are causing increased enrollment directly, through information dispersion about returns to schooling.

It may seem puzzling on its face that information is so localized given that people do travel for these jobs. Although it is beyond the scope of our data to prove this, it seems plausible that the type of people who actually have these jobs, and are therefore traveling for them, are not on the margin with respect to their childrens' school enrollment. It is likely to be parents and families who are less well off who are on this margin, and these may be individuals with more limited travel options and more limited exposure to others who work in these sectors. It seems plausible that, in fact, it is only having a physical center in a nearby location that provides information to this marginal group.

It is worth discussing the magnitude of our results, both in general and compared to other interventions to promote schooling in the developing world. Our preferred coefficient indicates that an additional ITES center prompts a 4.8% increase in school enrollment. Based on the National Family and Health Survey, a nationally representative survey run in 2005-2006, 84.4% of children aged 6-14 in our states attended school at any time in the previous year. Our coefficient implies that 25.6% of out-of-school children would be prompted to enroll by an ITES center.

Put differently, our estimates imply about a 4.1 percentage point increase in the enrollment rate. This number is comparable to enrollment effects of other interventions designed to increase schooling in the developing world. For example, the conditional cash transfers in PROGRESA increased schooling 3.4-3.6 percentage points (Schultz, 2004). A program in Kenya which provided school uniforms to girls in Kenya (worth about 1.75% of average yearly income) increased enrollment by 6 percentage points (Evans, Kremer and Ngatia, 2008). Miguel and Kremer (2004) found that administering deworming drugs decreased absence by 7 percentage points, although they do not report effects on enrollment. Our coefficient is similar to (although slightly smaller than) the 5.2 percentage point increase that Jensen (2012) identifies as a response to call center recruitment services for women.

From a policy standpoint, the results provide support for interventions which inform students about returns to schooling (as in Jensen, 2010 and 2012). In the absence of this type of policy, we would expect short-term gains in enrollment to be concentrated around areas with local ITES centers; the evidence in Section 6 suggests this concentration could be limited by broader information sharing.

Chapter 2

Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital[1]

2.1 Introduction

Human capital investment is an important determinant of economic growth (Mankiw *et al.*, 1992). However, there is still much debate over the drivers of human capital investment. The majority of empirical evidence from poor countries suggests the relationship is procyclical (see for example, Jacoby and Skoufias (1997); Jensen (2000); Thomas *et al.* (2004); Maccini and Yang (2009)). However, there is some evidence from Latin America suggesting countercyclical human capital investment (Duryea and Arends-Kuenning, 2003; Schady, 2004; Kruger, 2007).[2] Theoretically, the relationship is ambiguous; if time and income are important

[1]Co-authored with Manisha Shah

[2]All of these papers use school enrollment or years of schooling as their measure of human capital investment.

inputs into human capital, then increased wages could either increase or decrease human capital investment. As early as 1977, Rosenzweig and Evenson showed that higher wages are associated with *lower* schooling rates, due to increased opportunity costs of staying in school. If children react to higher wages by leaving school early to join the workforce, it could raise overall inequality in poor countries or even stunt long term growth.

We argue that at least part of the differences in these studies may be due to differential effects by age. In particular, if the opportunity cost of time for older children is affected by wages, then we would expect that the substitution effect would be more powerful for older children. In addition, if the human capital production function itself differs by age (for instance, if income-intensive inputs such as calories are more important at earlier ages), then we might also expect to see differential impacts of wage shocks by age. In this paper, we introduce a simple model of human capital investment from which we derive predictions about the effects of wages on human capital. Under certain conditions, our model predicts that increased wages during school years will negatively affect both schooling investment and overall human capital. In addition, the model predicts that in the presence of strong complementarity between early life consumption and later-life schooling investments, an increase in early life wages will positively affect both schooling investment and overall human capital.

We then estimate the comparative statics from this model, using rainfall fluctuations in rural India as quasi-random shocks to wages. We measure human capital using test scores from the ASER data from 2005-2009; we observe approximately 2 million rural children from almost every rural district in India. The data includes four distinct measures of literacy and numeracy for each child whether or not he is currently enrolled in school.[3] In addition, our data allow us to look at more standard educational measures such as school enrollment, drop out behavior, and being on track in school (age for grade). Since the survey is conducted every year over five years, we can control for age, year of survey, and district,

[3] This is rare since tests are primarily conducted at school, and thus scores are usually only available for currently enrolled kids who attended school on the day the test was given.

identifying off within-district variation in rain shock exposure.

We find that during years with positive rainfall shocks, school-age children (ages 5-16) score lower on simple math tests. When rainfall is higher, children are also less likely to attend school and are more likely to be working. In addition, children who experienced a positive rainfall shock in the previous year are more likely to have dropped out of school and less likely to be in the correct grade for their age.

We also estimate the impacts of early life rainfall shocks (in utero-age 4) on current test scores and schooling outcomes. We find that, by contrast, more early life rainfall is associated with higher test scores in both math and reading. In addition, children who experience positive rainfall shocks before age 5 are more likely to be enrolled in school and to be on track in school. Lastly, we investigate whether there are long-term impacts of these rainfall shocks on total years of schooling for adults aged 16-30 using a national labor and employment survey. We find that more rainfall during school years (particularly ages 11-13) lowers total years of schooling. This is also the age group where positive rainfall shocks significantly increase the likelihood of dropping out as these are the transition years from primary to secondary school so positive employment shocks are particulary detrimental to human capital investment during this period.

Our contribution to this literature is threefold. First, as far as we know, this is the first paper to document the possibility that positive productivity shocks can lead to lower levels of human capital attainment directly using test scores. Test scores are a much better measure of human capital as they measure output as opposed to the previous literature which has focused on school enrollment. Second, unlike the previous literature which focuses on shocks at certain critical ages in a child's development, we consider a child's entire lifecycle from in utero to age 16. This allows us to say something about the relative importance of time vs. income at all stages of a child's human capital development. We show that human capital investment is procyclical from the in utero phase to age three, but then becomes counter cyclical. Lastly, we provide new evidence on the long term effects of cumulative shocks on human capital attainment of young adults. While previous research has suggested that that these shocks represent simple intertemporal substitution of school time and that children make up these differences in human capital (Jacoby and Skoufias (1997); Funkhouser (1999)), we find quite the opposite. For example, children ages 11-13 complete approximately .2 more years for every drought experienced (and .2 fewer years for every positive rainfall shock relative to normal years). This constitutes a substantial shock to human capital attainment during a period when most children will already be on the margin between dropping out and continuing.

The findings from this paper are important from a policy perspective since workfare programs with guaranteed wages such as NREGA in India have become a popular means of redistribution as they provide aid to the poor along with corresponding work incentives.[4] However, workfare programs affect not only overall income, but also the prevailing wage and time cost of family members. It is possible such workfare programs could lead to decreased human capital production for certain individuals.

2.2 A Simple Model of Human Capital Investment

To fix ideas, we consider a simple model of human capital investment. Households consist of one child and one parent, and the parent maximizes the total utility of the household. The child lives for three periods. In the first period, the child is too young for school or work and only consumes. In the second period, the child also consumes, but in addition, she has one unit of time that can be spent either in school or working. In the third period, the household gets a payoff from the child's accumulated human capital.

Let c_t be consumption in period t where $t \in \{1, 2\}$, and $u_t(c_t)$ be the flow utility from consumption in period t, where $\frac{\partial u_t}{\partial c_t} > 0$ and $\frac{\partial^2 u_t}{\partial c_t^2} < 0$, $\forall t$. Let e_t be the human capital of the child in period t, and h be the human capital of the parent, which we assume does not change. Let $V(e_3)$ be the payoff to the household from the level of human capital in period

[4]Recent examples include programs in Malawi, Bangladesh, India, Philippines, Zambia, Ethiopia, Sri Lanka, Chile, Uganda, and Tanzania. However, the practice of imposing work requirements for welfare programs stretches back at least to the British Poor Law of 1834 (Imbert and Papp, 2012a).

3. β is the discount factor.[5] The total utility function of the household is thus

$$U(c_1, c_2, e_3) = u_1(c_1) + \beta u_2(c_2) + \beta^2 V(e_3)$$

Let $w_t \in (0, \bar{w})$ denote the wage in period t per unit of human capital, so that parents will be paid $w_t h$ and children will be paid $w_t e_t$ for each unit of time spent working in period t. w_t can be thought of as an aggregate productivity shifter, and in our empirical specifications will be proxied for by rainfall in agricultural areas. The wage is determined exogenously. In addition, let $s_2 \in [0, 1]$ denote the time that the child spends in school in period 2, and thus $(1 - s_2)$ will be the time she spends working. In the first period, household income will be earned entirely by the parent, and will be equal to his wage, $w_1 h$. In the second period, the household income will be equal to the earnings of the parent, $w_2 h$ plus the earnings of the child, $(1 - s_2)\, w_2 e_2$. We will abstract away from borrowing and savings decisions, so that consumption will always be equal to income in each period. Thus, consumption will be

$$c_1 = w_1 h$$

$$c_2 = w_2(h + (1 - s_2)e_2)$$

In the spirit of Cunha and Heckman (2007), we assume that human capital at date t is a function of human capital at date $t - 1$ plus any investments made in period $t - 1$. In this simple model, investments will take the form of either schooling or consumption. We will not allow for directed payments for human capital (such as books or tutors) or for parents to invest their own time to teach children. This is sensible in the context of rural India since primary school is free and compulsory,[6] and the Indian government has built many schools to keep the costs of attendance low.[7] In addition, the parents of these children often

[5]For convenient notation, we assume exponential discounting, even though in this model, the "periods" are of substantially different lengths. This has no effect on our results.

[6]While primary school is officially compulsory, in practice many children are in and out of school.

[7]For example, in 1971, 53 percent of villages had a public primary school, in 1991, 73 percent did (Banerjee

have very low human capital themselves, so it is unlikely that they are heavily involved in teaching their children literacy and/or numeracy.

In our three period model, human capital in period 1 is normalized to zero, and human capital in period 2 is only a function of the household's consumption in period 1, since the child is too young to attend school in this period. Human capital in period 3, however, will be a function of human capital in period 2, household consumption in period 2, and schooling in period 2. Thus, we have

$$
\begin{aligned}
e_1 &= 0 \\
e_2 &= f_2(c_1) \\
e_3 &= f_3\left(e_2, c_2, s_2\right)
\end{aligned}
$$

Without loss of generality, we will let $V(e_3) = e_3,\ \forall e_3$ for the remainder of the paper.[8] In addition, we will assume that $\frac{\partial f_3}{\partial e_2} \geq 0$, $\frac{\partial f_3}{\partial c_2} \geq 0$, $\frac{\partial f_3}{\partial s_2} \geq 0$, and $\frac{\partial f_2}{\partial c_1} \geq 0$. These are standard assumptions asserting that more schooling and consumption result in weakly more human capital. In addition, we will assume each input has diminishing marginal returns, that is, $\frac{\partial^2 f_3}{\partial e_2^2} \leq 0$, $\frac{\partial^2 f_3}{\partial c_2^2} \leq 0$, $\frac{\partial^2 f_3}{\partial s_2^2} \leq 0$, and $\frac{\partial^2 f_2}{\partial c_1^2} \leq 0$.

Since no choices are made in the first period, we can restrict our analysis to the decisions made starting in period 2. Thus, the parent solves

$$
\max_{s_2 \in [0,1]} \left\{ u_2\left(c_2\right) + \beta f_3\left(e_2, c_2, s_2\right) \right\} \text{ s.t. } c_2 \leq w_2\left(h + \left(1 - s_2\right) e_2\right)
$$

Since utility is increasing in consumption, and there is no borrowing or savings in this model, it will always be the case that $c_2 = w_2\left(h + \left(1 - s_2\right) e_2\right)$. Thus, we can substitute this into the maximization problem to get

and Somanathan, 2007), and today almost 100 percent of Indian villages have a primary school (Government of India, 2011).

[8] Because we allow for full flexibility of the human capital production function, we can make this simplification without loss of generality. However, it does change the interpretation of the function slightly, because it represents the household utility of human capital, rather than the productive capacity.

$$\max_{s_2 \in [0,1]} \left\{ u_2 \left(w_2 \left(h + (1 - s_2) e_2 \right) \right) + \beta f_3 \left(e_2, c_2, s_2 \right) \right\}$$

In addition, in order to ensure a globally concave objective function, and thus, a unique optimum, we will assume that

$$\left\{ \frac{\partial^2 u_2}{\partial c_2^2} + \beta \frac{\partial^2 f_3}{\partial c_2^2} \left(e_2, s_2, c_2 \right) \right\} \cdot \beta \frac{\partial^2 f_3}{\partial s_2^2} \left(e_2, s_2, c_2 \right) > \left\{ \beta \frac{\partial^2 f_3}{\partial c_2 \partial s_2} \right\}^2 \left(e_2, s_2, c_2 \right), \ \forall e_2, c_2, s_2$$

This assures that consumption and schooling are neither "too complementary" nor "too substitutable". That is, the absolute value of the cross partial with respect to consumption and schooling is smaller than that of the second derivatives. In addition, we will assume that

$$\lim_{s_2 \to 0^+} \frac{\partial f_3}{\partial s_2} = +\infty$$

and

$$\lim_{s_2 \to 1^-} \frac{\partial f_3}{\partial s_2} = 0$$

These assumptions, while not strictly necessary for analysis of a solution, will allow us to ignore corner solutions in which children spend either no time in school, or no time on productive work.[9] We want to focus on interior solutions because in practice, we find that most children in our data look like they are spending at least some time in school and some time on productive work.

At an interior optimum, parents will equalize the marginal utility consumption from forgoing school now with the marginal benefits of additional human capital later:

$$w_2 e_2 \frac{\partial u_2}{\partial c_2} = \beta \Theta(w_2, e_2, s_2^*, c_2^*)$$

[9] Because we have assumed that parents supply positive labor, and bounded schooling between 0 and 1, consumption will always be positive.

where

$$\Theta(w_2, e_2, s_2^*, c_2^*) = \frac{\partial f_3}{\partial s_2}(e_2, s_2^*, c_2^*) - w_2 e_2 \frac{\partial f_3}{\partial c_2}(e_2, s_2^*, c_2^*)$$

Households tradeoff the marginal benefit of additional utility from consumption with the net long-term benefit of schooling. Note that in an interior solution, since $\frac{\partial u_2}{\partial c_2},\ w_2,\ e_2 > 0$, it must be the case that $\Theta(w_2, e_2, s_2^*, c_2^*) > 0$. That is, at the optimum, schooling is a relatively better technology than working and consuming for turning time into human capital.

We are interested in the effect that wages have on the optimal level of schooling. That is, if wages increase, do children invest more or less in schooling? And, as a result, do overall levels of human capital increase or decrease? In this model, there are two relevant wages—those in early life and those during the child's school years. We will examine the effect of each of these wages on schooling choices and total human capital.

2.2.1 Effect of School-Aged Wages on Schooling and Human Capital

First, we examine the impact of second period wages, w_2 on the optimal choice of schooling, s_2^*, and the resulting level of human capital, e_3^*. From the first order condition,

$$\frac{\partial s_2^*}{\partial w_2} \propto \overbrace{-e_2\left(\frac{du_2}{dc_2} + \beta\frac{\partial f_3}{\partial c_2}\right)}^{\text{Substitution Effect (-)}} \overbrace{-\left(h + (1 - s_2^*)\, e_2\right)\frac{\partial^2 u_2}{\partial c_2^2}}^{\text{Income Effect (+)}} + \overbrace{\left(h + (1 - s_2^*)\, e_2\right)\beta\frac{\partial \Theta}{\partial c_2}}^{\text{Effect of } c_2 \text{ on Net Impact of Schooling}}$$

The effect of school-aged wages on the optimal level of schooling will depend on three things. First, increased wages increase the benefit to working, both through the utility in period 2, and through the benefit to human capital in period 3 (substitution effect). Second, increased wages will increase consumption, which will decrease the marginal utility of consumption (income effect). Third, the increase in consumption could affect the net impact of schooling. We think it is likely that this term is weakly positive. That is, as consumption increases, schooling becomes relatively better than consumption as a technology for turning time into human capital. If a child is starving, consumption is likely extremely important for the production of human capital. As their level of consumption increases, the benefits of consumption relative to schooling will likely decrease. Thus, even if income effects are small, if schooling becomes relatively more valuable as households get richer, we could still see schooling decrease when wages are higher. Which of these forces will dominate is an empirical question that we address in Section 4.1.

In addition, we can examine the impact of period 2 wages on period 3 human capital:

$$\frac{d}{dw_2}\left(f_3\left(e_2, c_2^*, s_2^*\right)\right) = \left(h + (1 - s_2^*)e_2\right)\frac{\partial f_3}{\partial c_2} + \Theta\frac{\partial s_2^*}{\partial w_2}$$

The first term in this expression is positive by assumption: it is the mechanical effect of higher wages on consumption, which in turn increases human capital. We also know that Θ is positive at any interior optimum, so if increased wages lead to increased schooling, we know that human capital will increase as a result. However, if increased wages *decrease* the optimal level of schooling, then the effect on human capital will be ambiguous. The sign will depend on whether this behavioral effect of lower investment will offset the mechanical increase in human capital due to consumption.

2.2.2 Effect of Early Life Wages on Schooling and Human Capital

In this model, the only way that early life wages affect the choice of schooling is through their effect on human capital in period 2. Because increased wages mechanically increase consumption in period 1, and human capital in period 2 is an increasing function of period 1 consumption, increased wages in period 1 will always result in increased human capital in period 2:

$$\frac{d}{dw_1}(e_2) = \frac{\partial f_2}{\partial c_1}\frac{\partial c_1}{\partial w_1} = h\frac{\partial f_2}{\partial c_1} > 0$$

Thus, in order to understand the effect of early life wages on schooling and later-life human capital, it is sufficient to study the effect of period 2 human capital on the optimal level of schooling and on period 3 human capital.

$$\frac{\partial s_2^*}{\partial w_1} = h\frac{\partial f_2}{\partial c_1}\frac{\partial s_2^*}{\partial e_2} \propto \frac{\partial s_2^*}{\partial e_2}$$

From the first order condition, we can derive the effect of period 2 human capital on the optimal choice of schooling:

$$\frac{\partial s_2^*}{\partial e_2} \propto \overbrace{-w_2\frac{\partial u_2}{\partial c_2}}^{\text{Substitution Effect (-)}} \quad \overbrace{-w_2^2 e_2\left(1-s_2^*\right)\frac{\partial^2 u_2}{\partial c_2^2}}^{\text{Income Effect (+)}} + \beta \quad \overbrace{\left[\frac{\partial \Theta}{\partial e_2} + w_2\left(1-s_2^*\right)\frac{\partial \Theta}{\partial c_2}\right]}^{\text{Net impact of additional } e_2 \text{ on } \Theta}$$

Increased period 2 human capital has three effects on the optimal level of schooling. First, increased human capital increases the value of work in the second period (substitution effect). Second, increased human capital leads to higher income, which reduces the marginal utility of consumption (income effect). Third, the net benefit of schooling, $\Theta(w_2, e_2, s_2^*, c_2^*)$, could be affected by the increase in human capital, in two ways. First, if there are "dynamic complementarities" in the sense of Cunha and Heckman (2007), we would expect the return to schooling to increase with early life investments. In addition, since additional early life human capital also increases consumption mechanically, this could also effect the net benefit of schooling even without dynamic complementarities. However, whether these effects will be large enough to overcome the substitution effects is again an empirical question, which we will address in Section 4.2.

In addition, we can examine the impact of childhood human capital on adult human capital. Again,

$$e_3^* = f_3\left(e_2, c_2^*, s_2^*\right)$$

$$\frac{d}{de_2}\left(e_3^*\right) = \frac{\partial f_3}{\partial e_2} + \frac{\partial f_3}{\partial c_2}w_2\left(1-s_2^*\right) + \Theta\frac{\partial s_2^*}{\partial e_2}$$

Intuitively, the first term can be thought of as the persistence of early life human capital, and is weakly positive by assumption. The second term takes into account the mechanical increase in consumption derived from an increase in early life human capital (through increased wages) and is also positive by assumption. As above, we know that Θ is positive at the optimum, so if early life human capital increases schooling, then it will unambiguously increase human capital as well. If not, it is not clear which effect will dominate.

Below, we outline a more restricted version of this model which makes clear predictions about the effect of wages at each age on human capital.

2.2.3 Example

Consider a simplified version of the model in which utility is linear in consumption

$$u_t(c_t) = c_t$$

and consumption in period 2 does not affect later-life human capital

$$e_3 = f_3(e_2, s_2)$$

While these assumptions are stark, they are not necessarily unreasonable starting points in the context of rural India. Since children earn a small percentage of total household income in practice, income effects from their earnings will likely be small. In addition, while much of the developmental literature focuses on the importance of nutrition and other consumption inputs at "critical periods" (generally from the in utero period until age 2) there is little evidence of the importance of nutrition at later ages. It is reasonable to think these effects might be second-order, relative to schooling and early life consumption.

In this case, the maximization problem is now

$$\max_{s_2 \in [0,1]} \left\{ \beta \left(w_2 \left(h + \left(1 - s_2 \right) e_2 \right) \right) + \beta^2 \left(f_3 \left(e_2, s_2 \right) \right) \right\}$$

and the interior optimum is characterized by the following first-order condition

$$w_2 e_2 = \beta \frac{\partial f_3}{\partial s_2}$$

Adding these restrictions, we can now make the following predictions:[10]

1. $\frac{\partial c_1}{\partial w_1} > 0$. Early life wages unambiguously increase child consumption in early life.

2. $\frac{d}{dw_2}\left(e_3^*\right) < 0$ and $\frac{\partial s_2^*}{\partial w_2} < 0$. School-aged wages decrease the optimal level of schooling *and* human capital.

3. $\frac{\partial s_2^*}{\partial e_2} > 0 \implies \frac{d}{de_2}(e_3^*) > 0$. If early life human capital increases schooling, then it also increases period 3 human capital.

4. $\frac{\partial s_2^*}{\partial e_2} > 0 \implies \frac{\partial^2 f_3}{\partial s_2 \partial e_2} > 0$. If early life human capital increases schooling, this implies that early life human capital increases also the returns to schooling (dynamic complementarities).

In the following sections, we will empirically estimate $\frac{\partial c_1}{\partial w_1}$, $\frac{\partial s_2^*}{\partial w_2}$, $\frac{\partial s_2^*}{\partial w_1}$, $\frac{d}{dw_1}\left(e_3^*\right)$, and $\frac{d}{dw_2}\left(e_3^*\right)$.

2.3 Background and Data

2.3.1 Cognitive Testing and Schooling Data

Every year since 2005, the NGO Pratham has implemented the Annual Status of Education Report (ASER), a survey on educational achievement of primary school children in India which reaches every rural district in the country.[11] We have data on children from 2005-2009, giving us a sample size of approximately 2 million rural children. The sample is a representative repeated cross section at the district level. The ASER data is unique in that its sample is extremely large and includes both in and out of school children. Since cognitive tests are usually administered in schools, data on test scores is necessarily limited to the sample of children who are enrolled in school (and present when the test is given). However, ASER tests children ages 5-16, who are currently enrolled, dropped out, or have

[10]For derivations, see Mathematical Appendix.

[11]This includes over 570 districts, 15,000 villages, 300,000 households and 700,000 children in a given year. For more information on ASER, see `http://www.asercentre.org/ngo-education-india.php?p=ASER+survey`

Table 2.1: *Summary Statistics*

ASER Summary Statistics (Ages 5-16)			
	Mean	Std. Dev.	Observations
Age	10.16	3.13	2,405,642
Math Score	2.62	1.31	2,104,110
Math Word Problem	1.26	.919	843,827
Reading Score	2.70	1.40	2,115,547
Dropped Out	.036	.188	2,193,040
On Track	.778	.415	1,919,939
Attendance	.863	.215	467,606
Never Enrolled	.028	.164	2,405,642
Rainfall Summary Statistics			
Rain Shock This Year	.148	.631	2,193,040
Rain Shock Last Year	.093	.631	2,193,040
Rain Shock in Utero	-.011	.572	2,405,642
Rain Shock at Age 1	-.024	.566	2,405,642
Rain Shock at Age 2	-.047	.558	2,405,642
Rain Shock at Age 3	-.058	.558	2,405,642
Rain Shock at Age 4	-.068	.561	2,405,642
NSS Sample			
Works (Ages 5–16)	.053	.224	305,065
Attends School (Ages 5–16)	.795	.403	303,244
ln Wages	5.85	0.91	167,041
Total Years of School (Ages 16–30)	6.21	4.92	306,925
Total Droughts (Ages -1–16)	3.25	1.12	306,925
Total Positive Shocks (Ages -1–16)	3.59	1.23	306,925

Notes: This table shows summary statistics from the ASER, NSS, and rainfall data.

never enrolled in school. In Table 2.1 we describe the characteristics of the children in our sample as well as their test scores.

The ASER surveyors ask each child four questions each in math and reading (in their native language). The four math questions are whether the child can recognize numbers 1-9, recognize numbers 10-99, subtract, and divide. The scores are coded as 1 if the child correctly answers the question, and 0 otherwise. In 2006 and 2007, children were also asked two subtraction word problems, which we use as a separate math score (Math word problem). The four literacy questions are whether the child can recognize letters, recognize words, read a paragraph, and read a story. We calculate a "math score" variable, which is the sum of the scores of the four numeracy questions. For example, if a child correctly recognizes numbers between 1-9 and 10-99, and correctly answers the subtraction question, but cannot correctly answer the division question, then that child's math score would be coded as 3.

The "reading score" variable is calculated in exactly the same way. In addition, the survey asks about current enrollment status and grade in school, and in 2008, attendance in the past week.[12] Table 2.1 summarizes the means of test scores and the schooling outcomes for children in the ASER sample.

2.3.2 Rainfall Data

To determine rainfall shock years and districts, we use monthly rainfall data which is collected by the University of Delaware.[13] The data covers all of India in the period between 1900-2008 and we use data from 1975-2008 in this paper. The data is gridded by longitude and latitude lines, so to match these to districts, we simply use the closest point on the grid to the center of the district, and assign that level of rainfall to the district for each year.

We define a positive shock as yearly rainfall above the 80th percentile and negative shock (drought) as rainfall below 20th percentile within the district. The "positive" and "negative" shocks should not be taken in an absolute sense—we are not comparing districts that are prone to higher rainfall to those that are prone to lower rainfall. These are simply high or low rainfall years for each district within the given time frame. For the analysis, we define "rain shock" as equal to 1 if rainfall is above the 80th percentile, -1 if rainfall is below the 20th percentile, and 0 otherwise. These are similar to the definitions employed in Kaur (2011) and Jayachandran (2006).[14] Figure 2.1 shows the prevalence of drought by district over time (for the years we have cohort variation in in utero drought exposure) and indicates there is both a lot of variation over time and across districts in terms of drought exposure. Between 6 and 48 percent of districts experience a drought in any given year, and 80 percent of the districts experience at least one drought in the 16 year period that we have child cohort

[12] More information on the ASER survey questions, sampling, and procedures can be found in the ASER data appendix.

[13] The data is available at: `http://climate.geog.udel.edu/~climate/html_pages/download.html#P2009`

[14] In previous versions of the paper we showed results separately for positive and negative rainfall shocks and using rainfall quintiles and the results are qualitatively similar.

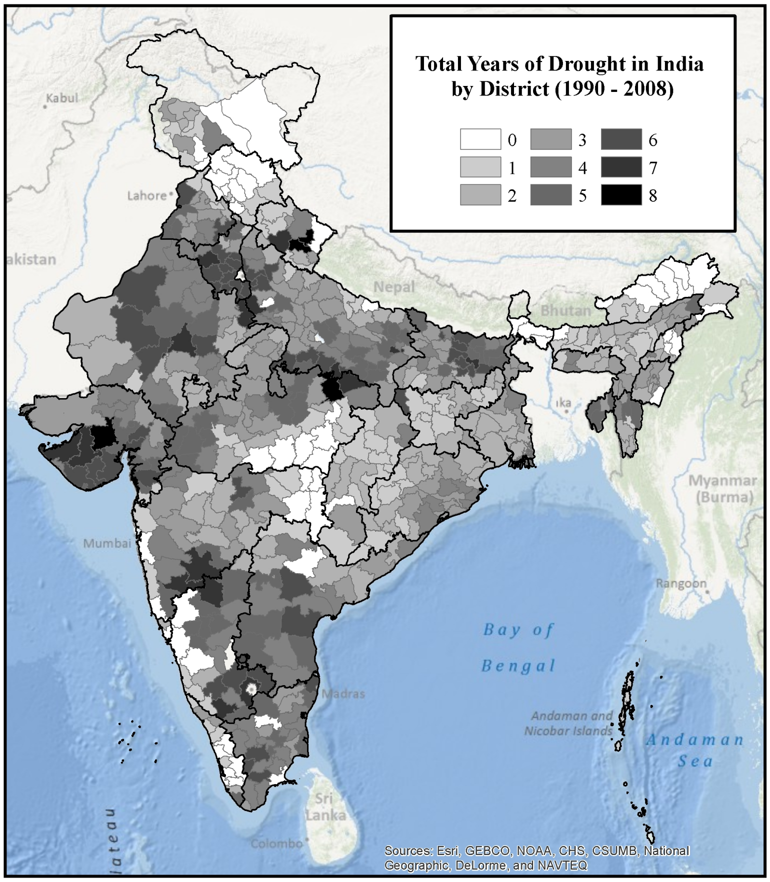


Figure 2.1: *Variation in Drought Across District and Time*

variation. Table B.11 shows the percent of districts each year that experience a drought or positive rainfall shock; the variation in rainfall across time and space is quite extensive.

In a data appendix, we explicitly test for serial correlation of rainfall because if droughts this year are correlated with droughts next year, it is difficult to tell the extent to which we are picking up the effects of a single shock or multiple years of rainfall shocks. However, we find no significant evidence of serial correlation across years. In addition, we check for spatial correlation. If there is significant within-district variation in rainfall, our district-level measure of rainfall variation might be missing the true effects for many of the children in our sample. However, we find that this type of very local variation is unlikely to be biasing our results (results available upon request).

2.3.3 Rainfall Shocks in India

In rural India, 66.2 percent of males and 81.6 percent of females report agriculture (as cultivators or laborers) as their principal economic activity (Mahajan and Gupta, 2011). Almost 70 percent of the total net area sown in India is rainfed; thus, in this context we would expect rainfall to be an important driver of productivity and wages. While there is plenty of evidence showing droughts adversely affect agricultural output and productivity in India (see for example Rao *et al.* (1988), Pathania (2007)), we also explore this question empirically using the World Bank India Agriculture and Climate Data set. In Table B.10 we show results from regressions of rice, wheat, and jowar yields on rainfall shocks. In drought years, crop yields are significantly lower regardless of the type of crop (and the opposite is true in positive rain shock years). In Table 2.2 we will test explicitly for rainfall's effect on wages for both adults and children in rural India.

2.3.4 NSS Data

To examine the impact of drought on work and wages, we use the NSS (National Sample Survey) Rounds 60, 61, 62, and 64 of the NSS data which was collected between 2004 and 2008 by the Government of India's Ministry of Statistics. This is a national labor and employment survey collected at the household level all over India. This dataset gives us measures of employment status as well as wages at the individual level. Given the potential measurement error in the valuation of in-kind wages, we define wages paid in money terms. We use data from all rural households in this survey and merge with our district level rainfall data to explore the relationship between weather shocks, labor force participation, school attendance, and wages.

2.4 Empirical Strategy and Results

In Section 2.2, we outlined a model in which the effects of early life and school-aged wages on human capital was ambiguous. However, when we assume that utility is linear in consumption and school-aged consumption does not impact later-life human capital, the theory in Section 2.2.3 predicts that school-aged wages will negatively impact human capital investment, and that early life wages will increase schooling if and only if there are large enough dynamic complementarities in the human capital production function. We now estimate these comparative statics using the test score and schooling data from India.

To estimate the impact of school-aged wages on schooling and human capital, we estimate the impact of current year rainfall shocks on current levels of schooling and human capital. To determine the effects of early life wages on human capital outcomes, we need to regress current test scores on lagged rainfall since we do not have measures of human capital for very young children. In both cases, we will be relying on the quasi-random nature of droughts and positive rainfall shocks within districts as a natural shifter of rural wages. We outline both strategies in detail below.

Before we move to the reduced form estimation of the effect of rainfall shocks on wages, we first need to show that rainfall and agricultural productivity (and thus, wages) have a positive relationship. While there is extensive literature in economics and other fields both documenting this fact and using it to estimate economic parameters of interest (see for example Jayachandran (2006); Maccini and Yang (2009); Jensen (2000); Kaur (2011)), we also test for the relationship using our data.

In Table 2.2, we measure the effect of rainfall shocks on wages for children ages 5-16, as well as adult men and women using NSS data. We find that for all three groups, positive rainfall shocks result in increased wages. Children's and women's wages are more responsive to rainfall shocks than men's wages. In Appendix Table B.10, we also show that agricultural yields are significantly higher across all types of crops in years with more rainfall, controlling for labor and other inputs. These results give us confidence that rainfall shocks are indeed a productivity, and thus, wage shifter in this context.

Table 2.2: *Effect of Rain Shocks on Wages*

Dependent Variable:	*ln Wages (Ages 5-16)*	*ln Wages (Female adults)*	*ln Wages (Male adults)*
Rain Shock This Year	.19	.12	.06
	(.03)***	(.02)***	(.01)***
Observations	6894	43,669	116,478
Mean Dependent Variable	5.34	5.41	6.04

Source: Wages data from rounds 60, 61, 62, and 64 of NSS. Rainfall data from University of Delaware. Notes: This table shows estimates of the effect of rain shocks on ln wages using OLS regressions. All regressions contain district fixed effects and control for age. Column 1 additionally controls for sex. We restrict column 1 to children ages 5-16, column 2 to adult women, and column 3 to adult men. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

2.4.1 Contemporaneous Rainfall Regressions

Theory predicts that if the effect of school-aged consumption is small, then we should expect to see that wages during the school years (w_2), are associated with lower levels of schooling and lower overall human capital (that is, $\frac{d}{dw_2}\left(e_3^*\right)$ and $\frac{\partial s_2^*}{\partial w_2}$ are negative). To test this empirically, we estimate the regression:

$$S_{ijty} = \alpha + \beta_1\delta_{j,y} + \beta_2\delta_{j,y-1} + \zeta\theta_{j,t} + \gamma_j + \phi_t + \psi_y + \epsilon_{ijty} \tag{2.1}$$

where S_{ijty} is the measure of human capital or schooling for student i in district j born in year t and surveyed in year y. As measures of e_3, we use math and reading test scores, as well as "on track" which is a measure of age-for-grade. We define on track as a binary variable which indicates if a child is in the "correct" grade for his/her age. The variable is coded 1 if age minus grade is at most six. That is, if an eight year old is in second or third grade, he is coded as on track, but if he is in first grade, he is not. We use self-reported attendance and an indicator of having dropped out of school as two measures of s_2, schooling in period 2. $\delta_{j,y}$ is rain shock in district j in year y and $\delta_{j,y-1}$ is a lagged rain shock. β_1 is the impact of current year rain shock on the various cognitive test scores and schooling outcomes. We also control for early life rainfall exposure by including $\theta_{j,t}$, a vector of early life rainfall shocks from in utero to age 4. γ_j is a vector of district fixed effects, ϕ_t is a vector of age fixed effects, and ψ_y is a vector of year of survey fixed effects. This specification allows us to compare children who are surveyed in different years from the same district. Since our regressions contain district level fixed effects, the coefficient will not be biased by systematic differences across districts. Standard errors are clustered at the district level.

In Panel A of Table 2.3 we report the results from Equation 1 estimating the impact of contemporaneous rainfall shocks on test scores and schooling outcomes of children ages 5-16.[15] The coefficient on math score is -.02, which means that, relative to a positive rainfall year, children tested in a drought year score .04 points better (or 1.5 percent) on the math test. The coefficient on math word problem is -.05 which means that relative to a positive rainfall year, children in a drought district score 0.1 points more (or 8 percent). While rain shocks this year do not impact reading scores, rainfall shocks last year significantly decrease these scores as well.

While rainfall shocks in the current year have little effect on age for grade (i.e. being on track) or dropping out, rainfall shocks in the previous year affect both age for grade and dropping out significantly. This makes sense since being a drop out or being held back are variables that are likely more affected by previous behavior than behavior in the current year. Children in a positive rainfall shock year are .4 percentage points more likely to report having dropped out in the following year, relative to children tested in drought years (this is an increase of 10% from a mean of .036). Likewise, children tested in a positive shock year are 2 percentage points less likely to be on track, relative to a drought year. In addition, children who experience a current drought are 4 percentage points more likely to have attended school in the previous week (from a mean of 86 percent) relative to a positive rainfall shock.

In Panels B and C of Table 2.3, we report these coefficients separately estimated for children ages 5-10 and ages 11-16. Most of the magnitudes are similar in size, although the effect of rainfall on dropouts and being on track appears to be almost entirely driven by older children. Indeed, Figure **??** shows the the coefficient of lagged rain shock on dropping

[15] We can only use ASER rounds 2005-2008 for Table 2.3 (not the 2009 round) because the rainfall data is only available to 2008 so there is no measure of rain shock this year for children in the 2009 ASER round.

Table 2.3: *Effect of Contemporaneous Rainfall Shocks on Human Capital*

	Dependent Variable:					
	Math Score	Math Word Problem	Read Score	Dropped Out	On Track	Attendance
			Panel A: Ages 5-16			
Rain Shock This Year	-.02	-.05	.002	.0002	.002	-.02
	(.01)*	(.02)***	(.01)	(.0008)	(.002)	(.006)***
Rain Shock Last Year	-.02	-.04	-.02	.002	-.01	
	(.01)	(.02)*	(.01)*	(.0009)**	(.003)***	
Observations	2,104,110	843,827	2,115,547	2,193,040	1,919,939	467,606
Mean Dependent Variable	2.62	1.26	2.70	.036	.778	.863
			Panel B: Ages 5-10			
Rain Shock This Year	-.02	-.07	-.001	.0006	.006	-.02
	(.01)	(.03)***	(.01)	(.0004)	(.002)**	(.006)***
Rain Shock Last Year	-.01	-.02	-.02	.0008	-.006	
	(.01)	(.03)	(.01)	(.0004)*	(.003)**	
Observations	1,154,292	383,271	1,162,482	1,189,704	914,129	254,770
Mean Dependent Variable	2.06	.785	2.08	.009	.877	.843
			Panel C: Ages 11-16			
Rain Shock This Year	-.01	-.04	.005	-.0004	-.001	-.014
	(.01)	(.02)**	(.009)	(.001)	(.003)	(.006)**
Rain Shock Last Year	-.02	-.05	-.03	.004	-.02	
	(.01)*	(.02)***	(.01)**	(.002)**	(.004)***	
Observations	949,818	460,556	953,065	1,003,336	1,005,810	212,836
Mean Dependent Variable	3.31	1.65	3.46	.070	.688	.887

Source: Test Score Data from ASER 2005-2008. Rainfall Data from University of Delaware. Notes: This table shows estimates of β_1 β_2 and ζ from equation (1). "Math Score" and "Read Score" range from 0-4. "Math Word Problem" ranges from 0-2 and was only available in 2006 and 2007. "On Track" is equal to one if age minus grade is at least six, and zero otherwise. All regressions also control for in utero to age 4 rainfall shocks. Columns 1-5 contain fixed effects for district, year and age. Since attendance is only observed in 2008, column 6 contains fixed effects for state and age. Panel A includes entire ASER sample ages 5-16 years old, Panel B restricts to ages 5-10; and Panel C to ages 11-16. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

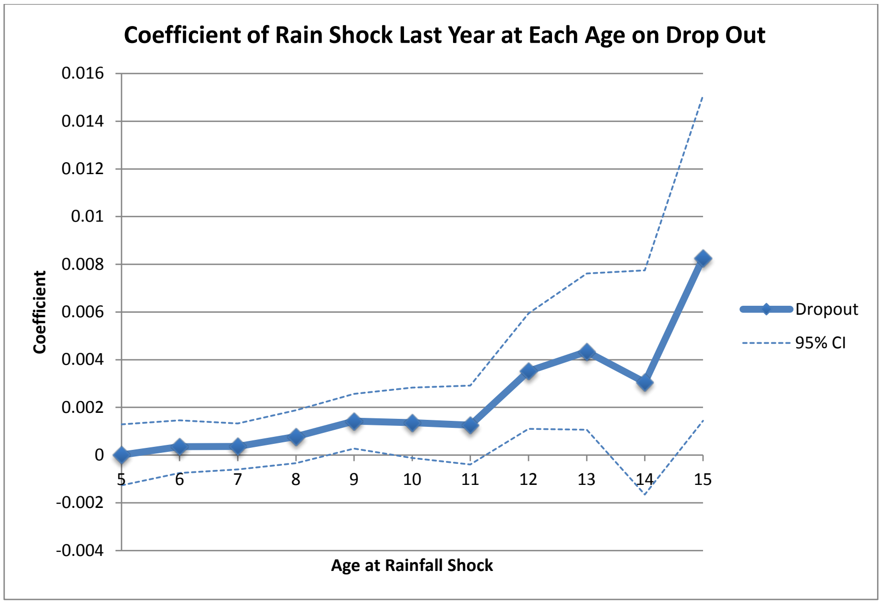


Figure 2.2: *Effect of Rainfall Shocks on Dropout Rates, by Age*

out estimated for each age separately. It appears that experiencing a positive rainfall shock from age 12 onward results in a higher likelihood of dropping out, though the estimates are noisy. This makes sense since this is the period children transition from primary to secondary school and when outside job opportunities during high rainfall years might lure them away from school.

In Table 2.4 Panel A, we also estimate the impact of rain shocks on children's reported "primary activity" using NSS data to corroborate the ASER attendance results. We find that during positive rainfall shocks, children are 2.5 percent less likely to report attending school and 20 percent more likely to report working relative to a drought year. Interestingly, the attendance results in Tables 2.3 and 2.4 are similar across both the datasets. In Panels B and C, we report these coefficients separately estimated for children ages 5-10 and ages 11-16. As in the ASER data, the effects are much larger for older children. Note that these

Table 2.4: *Effect of Rain Shocks on Schooling and Child Labor*

	Dependent Variable:	
	Attends School	*Works*
	Panel A: Ages 5-16	
Rain Shock This Year	-.01	.005
	(.003)***	(.001)***
Rain Shock Last Year	-.004	.001
	(.003)	(.001)
Observations	297,470	299,271
Mean Dependent Variable	.795	.053
	Panel B: Ages 5-10	
Rain Shock This Year	-.005	.002
	(.004)	(.0006)***
Rain Shock Last Year	-.002	.001
	(.004)	(.0008)
Observations	154,291	155,274
Mean Dependent Variable	.821	.007
	Panel C: Ages 11-16	
Rain Shock This Year	-.02	.008
	(.003)***	(.002)***
Rain Shock Last Year	-.008	.002
	(.003)**	(.002)
Observations	143,179	143,997
Mean Dependent Variable	.766	.103

Source: Attends school and child labor from rounds 60, 61, 62, and 64 of NSS. Rainfall data from University of Delaware. Notes: This table shows estimates of β_1 and β_2 from equation (1). The coefficients represent the effect of rain shocks on school attendance and working. Panel A restricts the sample to children ages 5-16, Panel B to ages 5-10; and Panel C to ages 11-16. All regressions contain district fixed effects and control for age and sex. All columns contain controls for early life rainfall shock exposure (in utero-age 4). Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

categories (child primarily attends school or primarily works) are mutually exclusive in the questionnaire, so that any intensive margin changes in work or attendance are not picked up here. Because of this, it is possible that these results understate the rain dependent substitution between schooling and labor for children.

We find that both schooling and human capital are lower during higher rainfall years when children are over the age of 5. These results are consistent with predictions of the the simplified model in Section 2.2.3.

2.4.2 Early Life Rainfall Exposure

We estimate the effect of early life wages on human capital for two reasons. First, our theory predicts that if schooling is increased by early life wages, then human capital will increase as well. This prediction is directly testable in the data. In addition, if we do find that early life wages increase schooling investment ($\frac{\partial s_2^*}{\partial w_1} > 0$), then this is evidence for dynamic complementarities in the human capital production function.

We use a lagged rainfall specification to estimate the effect of early life wages on later schooling ($\frac{\partial s_2^*}{\partial w_1}$) and human capital ($\frac{d}{dw_1}(e_3^*)$) and to investigate longer-term effects of both early life and school-aged wages on adult human capital ($\frac{d}{dw_1}(e_3^*)$ and $\frac{d}{dw_2}(e_3^*)$). In all specifications, we look at lagged effects of rainfall shocks on current outcomes exploiting cohort variation in rain exposure.[16]

To examine the effect of early life wages on human capital and schooling, we estimate the following regression:

$$S_{ijhty} = \alpha + \zeta\theta_{j,t} + \lambda_h + \phi_t + \psi_y + \epsilon_{ijhty} \tag{2.2}$$

where S_{ijhty} is the measure of human capital or schooling of student i in district j born in year t and surveyed in year y, who is a member of household h. Again we use math and reading scores and "on track" as our measures of e_3 and "never enrolled in school" as a measure of s_2. $\theta_{j,t}$ is a vector of early life rain shocks from in utero to age 4, λ_h is a vector of household fixed effects, ϕ_t is a vector of age fixed effects, and ψ_y is a vector of year of survey fixed effects. ζ is the vector of coefficients of interest and it is the impact of early life rainfall shocks at each age on human capital outcomes. Comparing children from the same district who were born in different cohorts allows us to use household fixed effects in this regression.[17] Household fixed effects allow us to rule out the possibility that the results

[16]In our data, we do not observe exact date of birth, only age at time of survey. We generate year of birth=survey year-current age; but this measure of rainfall at each age will be somewhat noisy. We examine this issue in detail in an appendix and show that the main results are similar when we correct for measurement error.

[17]If drought exposure is indeed IID, and there are no intervening mechanisms which could affect outcomes,

are driven by lower ability children showing up more frequently in drought cohorts due to selective migration and/or fertility. Standard errors are clustered at the district level. We discuss potential selection issues in Section 2.5 below.

Table 2.5 presents the main estimates of the effect of early life rainfall on test scores and schooling outcomes. In the first three columns, we examine the effect of rainfall on math test scores, math word problems, and reading test scores. The coefficient on rain shock between the in utero period and age 3 ranges from .01-.02, which implies that for each year of exposure to positive rainfall, children score .02-.04 points higher on these tests relative to drought years, and for each year of exposure to drought, they score .02-.04 points lower relative to a positive shock year. In column 4, we show that drought exposure at every year from the in utero period to age 4 is associated with a higher probability of the child never having enrolled in school. The coefficients range from -.002 to -.003, relative to a mean of .028. In column 5, we show that from the in utero period to age 2, exposure to positive rainfall shocks significantly increases the probability of a child being on track. The coefficients range from .01-.02, from a mean of 0.781. These results are consistent with the idea that both schooling investments and human capital achievement are higher when wages are higher in early life.

Additionally, our model predicts that children's early life consumption should increase with early life wages ($\frac{\partial c_1}{\partial w_1} > 0$) under a wide range of assumptions. We test this prediction in Table 2.6 using IHDS 2004–2005 data for children ages 1-5.[18] We regress weight for age z-scores (using the 2006 WHO child growth standards for children ages 1-5) on rainfall shocks. We show that children have significantly lower weight for age z-scores in drought years (by .12 standard deviation) and higher weight for age z-scores in positive rainfall

this specification should yield exactly the same results as using district fixed effects, except that it is identified off of households with more than one child. However, it is possible that parents could react to one child's drought exposure by reallocating resources within the household, either by shifting them toward or away from the affected child. Thus, other children in the household could be affected by their sibling's drought exposure. Regressions estimated with district fixed effects are qualitatively similar, and available upon request.

[18] The India Human Development Survey (IHDS) is a nationally representative survey of 41,554 households in 1503 villages and 971 urban neighborhoods across India. The data and more information is available online at ihds.umd.edu.

Table 2.5: *Effect of Early Life Rainfall Shocks on Human Capital*

	Dependent Variable:				
	Math Score	Math Word Problem	Read Score	Never Enrolled	On Track
Rain Shock In Utero	.01	.006	.01	-.002	.02
	(.004)***	(.004)	(.004)***	(.0004)***	(.002)***
Rain Shock Year of Birth	.01	.009	.01	-.002	.02
	(.004)***	(.004)**	(.004)**	(.0004)***	(.002)***
Rain Shock at Age 1	.01	.02	.01	-.003	.02
	(.004)***	(.005)***	(.004)***	(.0004)***	(.002)***
Rain Shock at Age 2	.01	.02	.01	-.003	.01
	(.004)**	(.004)***	(.004)***	(.0004)***	(.002)***
Rain Shock at Age 3	.001	.008	.007	-.002	.003
	(.004)	(.005)*	(.004)	(.0004)***	(.002)
Rain Shock at Age 4	.002	-.008	.01	-.002	.005
	(.004)	(.004)*	(.005)***	(.0004)***	(.002)***
Observations	2,350,976	843,827	2,362,940	2,405,642	2,100,717
Mean Dependent Variable	2.62	1.26	2.71	.028	.781

Source: Test Score Data from ASER 2005-2009. Rainfall Data from University of Delaware. Notes: This table shows estimates of ζ from equation (2), the effect of early life rainfall shocks on current test scores and schooling outcomes. "Math Score" and "Read Score" range from 0-4. "Math Word Problem" ranges from 0-2 and was only asked in 2006 and 2007. "On Track" is equal to one if age minus grade is at least six, and zero otherwise. All regressions contain fixed effects for household, year and age. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

shock years. Consistent with our model, we find evidence that early life consumption is higher when rainfall levels are higher.

Though others have examined the impact of early life shocks on health outcomes, wages, and total years of schooling, there is little medium term evidence on human capital directly (i.e. test scores). Our results are similar to Akresh *et al.* (2010) who also find negative effects of shocks in utero and infancy and Maccini and Yang (2009) who find positive effects of early life rainfall on human capital. However, both of these papers find different effects for different groups and ages. Akresh *et al.* (2010) find that the most important year is the in utero year while Maccini and Yang (2009) find it is the year after birth (and only for girls). We find largely similar effects for children under three and do not find large differences by gender. Our coefficients suggest that the in utero effects are slightly larger for girls and that girls exposed to droughts are more likely to not enrol in school relative to boys, but standard errors in most cases do not allow us to detect significant differences between boys and girls (results by gender are shown in Appendix Tables B.13-B.16).

Table 2.6: *Effect of Rainfall Shocks on Child Weight*

	Weight for age z-score
Rain Shock This Year	.12
	(.05)**
Rain Shock Last Year	.22
	(.06)***
Observations	15,307
Mean Dependent Variable	-1.516

Source: Data on child weight from IDHS in 2004-2005. Rainfall Data from University of Delaware. Notes: This table shows our estimates of the effect of rainfall shocks on weight for age z-scores for children ages 1–5 (or $\frac{\partial c_1}{\partial w_1}$). These are anthropometric z-scores using the 2006 WHO child growth standards. The regression contains age, year, and state fixed effects. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

We find that early life rainfall is associated with both higher early life consumption, and also higher schooling investments and levels of human capital in later childhood. These results are consistent with the predictions of the simple model in Section 2.2.3 in the presence of strong dynamic complementarities in the human capital production function.

2.4.3 Long Term Effects of Rainfall Shocks

We are also interested in the effect of total childhood rainfall shocks experienced on adult human capital ($\frac{d}{dw_1}\left(e_3^*\right)$ and $\frac{d}{dw_2}\left(e_3^*\right)$). Table 2.3 indicates that students in districts with positive rainfall shocks have lower contemporaneous test scores. It is possible, however, that this represents simple intertemporal substitution of school time, and that children make up these differences in human capital over time. In fact, this is what the empirical literature to date suggests (see Jacoby and Skoufias (1997); Funkhouser (1999)). Table 2.3 suggests that there are lagged effects for positive rainfall shocks, perhaps due to the increased propensity to drop out in these years as well.

To test for this, we estimate Equation 2 using the NSS data on adults (ages 16-30 since rainfall goes back to 1975). However, instead of using only early life exposure, we replace $\theta_{j,t}$ with a vector of rain shocks from the in utero period to age 16. Our outcome variable for this specification is total years of schooling. We also use district, rather than household, fixed effects in this specification.[19]

Table 2.7 indicates that starting at approximately age 3, in almost every year of life, higher rainfall is associated with lower levels of schooling. The magnitudes are largest between ages 11-13 (a positive rainfall shock at age 12 reduces total years of schooling by approximately .23 years relative to a normal rainfall year). This makes sense, since the transition from primary to secondary school is a common time for students to drop out of school. We graph the coefficients from this regression in Figure 2.3. The results clearly indicate that the worst time to experience a positive rainfall shock for total years of schooling is in these transition years from primary to secondary. This is already when many children drop out of school as shown in the ASER data and experiencing a positive rainfall shock exacerbates this problem.

We find evidence in this section that the effects of rainfall on schooling and human capital can last into adulthood. Those who experienced higher rainfall on average in later childhood have fewer total years of schooling as adults. Thus, it is likely that students are not substituting across time, but that these changes in human capital represent real, lasting differences.

2.5 Alternative Explanations

Since we use rainfall shocks as a proxy for wages in this paper, other aspects of abnormally high or low rainfall that affect human capital could be a threat to our identification. We discuss three such possibilities in this section. First, we examine whether direct disease mechanisms, caused by excess water from high rainfall years, could cause children to become sick and attend school less. Second, we explore whether school lunches, now a common phenomenon in India, could be driving children to attend school more during drought years. Third, we examine whether the rain shocks could affect the outside options for teachers, affecting the quality of schooling directly. Each of these explanations could, in

[19] This is because the NSS data is a household survey of adults, and does not include information on the siblings of all adults in the household.

Table 2.7: *Effect of Rain Shocks on Total Schooling*

Dependent Variable:	Years of Education (Ages 16-30)
In Utero Rain Shock	.006 (.02)
Rain Shock in Year of Birth	.001 (.02)
Rain Shock at Age 1	-.05 (.02)**
Rain Shock at Age 2	-.03 (.03)
Rain Shock at Age 3	-.11 (.03)***
Rain Shock at Age 4	-.09 (.03)***
Rain Shock at Age 5	-.10 (.03)***
Rain Shock at Age 6	-.06 (.03)**
Rain Shock at Age 7	-.02 (.03)
Rain Shock at Age 8	-.05 (.03)*
Rain Shock at Age 9	-.09 (.03)***
Rain Shock at Age 10	-.11 (.03)***
Rain Shock at Age 11	-.27 (.03)***
Rain Shock at Age 12	-.23 (.03)***
Rain Shock at Age 13	-.27 (.03)***
Rain Shock at Age 14	-.08 (.03)***
Rain Shock at Age 15	-.13 (.02)***
Rain Shock at Age 16	-.09 (.03)***
Mean Dependent Variable	6.04
Observations	306,925

Source: Years of schooling from rounds 60, 61, 62, and 64 of the NSS. Rainfall data from University of Delaware. Notes: This table shows our estimates of $\frac{\partial s_2}{\partial w_1}$ and $\frac{\partial s_2}{\partial w_2}$, the effect of childhood rain shocks on total years of schooling using data for individuals 16-30. The regressions contain district fixed effects and control for age and sex. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

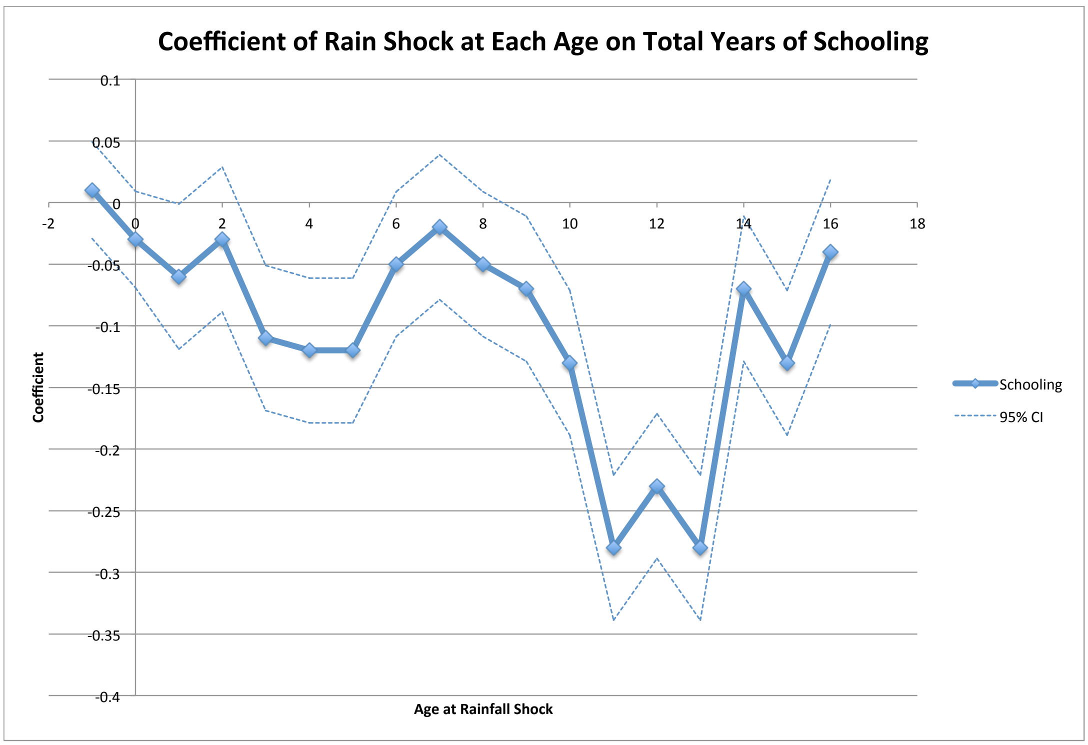


Figure 2.3: *Effect of Rainfall Shocks on Total Years of Schooling*

theory, bias our estimated contemporaneous coefficients from Table 3 upward. Below, we examine each of these explanations in turn, and find evidence in each instance that they are unlikely to be driving our results. We then explore how selective migration, mortality, and/or fertility responses may impact our main results.

2.5.1 Healthier Children

If less rainfall leads to lower endemicity of particular diseases, this could cause children to attend school more during drought years for reasons unrelated to their outside option. Two common diseases for children in India for which there has been a link discussed between weather patterns and disease rates are diarrhea and malaria. Rainfall variability as manifest through more frequent flooding has been linked to increases in the prevalence of diarrhea in studies in India, Bangladesh, Mozambique, and even in the USA (Curriero *et al.*, 2001; IPCC, 2007). However, other studies have shown that shortage of rainfall in the dry season increases the prevalence of diarrhea (see for example Sub-Saharan Africa (Bandyopadhyay *et al.*, 2012)). In fact, heavy rainfall events decreased diarrhea incidence following wet periods in Ecuador (Carlton *et al.*, 2013).

The evidence for malaria is similarly controversial. While we generally think more rain is associated with higher rates of malaria, there is evidence that droughts result in river margins retreating leaving numerous pools suitable for vector breeding exacerbating the spread of malaria (Haque *et al.*, 2010). Nevertheless, since malaria prevalence varies considerably by region, we can test for the possibility that differences in malaria infections during drought years might explain the test score results. In Table B.18 we re-estimate our contemporaneous shock regressions including an interaction of rainfall shock with an indicator for whether the district is in a high-malaria state (i.e. Orissa, Chhattisgarh, West Bengal, Jharkhand, and Karnataka (Kumar *et al.*, 2007)). The results in Table B.18 indicate that there is no additional statistically significant effect of rainfall shocks in malaria states, and thus it is unlikely this channel is driving the contemporaneous test score results.

We test for the overall health impacts of rainfall shocks on children ages 5-16 using the

IHDS data in Table B.17. The concern is that for whatever reason, children are healthier during drought years which results in them attending school more and doing better on their tests. In column 1 we regress the number of days ill in the past month due to diarrhea, cough, and/or fever. The results indicate that children are actually healthier in positive rainfall shock areas. Children spend 0.52 fewer days (or 10 percent) being ill. In column 2, we regress ln health expenditures (doctors, medicine, hospital and transportation) on rainfall shocks. Again the results suggest that children are healthier in positive rainfall shock years. Medical health expenditures are 44 percent lower in positive rainfall shock years, relative to drought. This is despite the fact that incomes are *higher* in positive shock years and *lower* in negative shock years. Therefore, we can conclude that children do not appear to be healthier in drought years in this context.

2.5.2 School Lunches

In November 2001, in a landmark reform, the Supreme Court of India directed the Government of India to provide cooked midday meals in all government primary schools (Singh *et al.*, forthcoming). Since that time, many schools have begun lunch programs, but compliance is still under 100 percent. One concern is that schools might be more likely to serve lunches during droughts and that students and parents respond to this by sending their children to school for the meals. We test whether schools are more likely to serve lunches during droughts using the ASER School Survey data, and do not find any evidence of this. In fact, column 2 of Table B.19 indicates that lunches are *more* likely to be provided in positive rainfall shock years. This makes sense since these are the years everyone is better off so districts and/or schools may have more resources to provide lunches.

2.5.3 Teacher Attendance

Tables 2.4 and 2.2 illustrate that employment and wages are affected by rainfall shocks. Thus, as the outside option for students and parents increases in value, so does the outside option for teachers. It is possible that the effects of rainfall shocks on test scores, and even on student absence and dropout rates, could be the result of teacher absences. We think this is unlikely in the context of India, because while absence rates for teachers are high overall (Chaudhury *et al.*, 2006), teachers are well-educated and well-paid workers, and the wages that are most affected by rainfall shocks are those for agricultural laborers who earn very little. The additional wage income available during good years for day labor such as weeding and harvesting is small relative to teacher's salaries.[20]

In column 1 of Table B.19 we show the impact of rainfall shocks on teacher absence rates recorded by surveyors in the ASER School Survey. The results indicate that teachers are less likely to be absent from school in positive rainfall shock years. Therefore, teacher absence cannot be the main driver of the contemporaneous test score results.[21]

2.5.4 Selective Migration in Contemporaneous Regressions

The primary selection concern for our main results in Table 2.3 is that ASER is sampling a different set of children in districts experiencing higher than average rainfall relative to districts experiencing lower rainfall. Specifically, if higher ability children are systematically less likely to be surveyed when rainfall is highest, this could bias our results upward. Fortunately, ASER has a procedure designed to reduce sample selection as much as possible. Enumerators are instructed to visit a random sample of households only when children are likely to be at home; they must go on Sundays when children are not in school and no one works. If all children are not home on the first visit, they are instructed to revisit once they are done surveying the other households (ASER, 2010).

This would not alleviate the issue if these students are leaving their districts permanently when rainfall is particularly high (or low). However, migration rates in rural India are extremely low. For example, Topalova (2005) using data from the National Sample

[20]Indeed, wages in the educational sector can be as much as 10 times higher than wages in the agricultural sector (NSS 2005 data).

[21]It is important to note that the school lunch and teacher absence results presented in Table B.19 are suggestive because the schools sampled in the ASER School Survey (unlike the households) are not a representative, random sample of schools in the district.

Surveys finds that only 3.6 percent of the rural population in 1999-2000 reported changing districts in the previous 10 years. Munshi and Rosenzweig (2009) using the Rural Economic Development Survey also conclude that rural emigration rates are low. Indian Census data from 2001 shows that the inter-district rural migration rate for all ages is .078. However, the rates drops to .02 when we look at children ages 5-14. Interestingly, the main reason for migration for females is marriage (65 percent of female migrants) and work/employment for men (37.6 percent of male migrants).

In Table B.21 using NSS data from round 55 (1999-2000) we regress whether members of households have stayed in the same village for the past 6 months or more on rainfall shocks. This allows us to test whether individuals are responding to positive and/or negative rainfall shocks by migrating. In columns 2 and 4 we restrict our samples to children ages 5-16, the same ages as the ASER sample. The results are very much in line with the census data. Firstly, only about 2 percent of rural households report to having moved in the last 6 months (or more). However, it does not appear that migration decisions are being driven by rainfall shocks. The magnitudes of the coefficients are close to 0 and the results are not statistically significant.

We can take these coefficients seriously and bound our results in the spirit of Manski (1990). We assume the "worst case scenario" for our hypothesis: that all excess movement into drought districts is high-scoring children, and all movement into positive shock districts is from low-scoring children. Essentially, we want to ask whether there is *any* way the amount of rain-responsive migration could be driving our results. In simulations, we find that even under the starkest assumptions (that all children who move into a drought district scored 4 on all tests, and all children who moved into positive shock districts scored 0 on all tests), our results are remarkably unchanged. 95% of the simulation results changed the coefficients for math score, math word score, and reading score by less than .0007, .0003, and .0006 respectively. Migration rates, particularly short-term migration rates among young children, are simply too small to explain our results.

Lastly, we are encouraged by the fact that the NSS results tell the same story as the ASER test score results. For the NSS survey, children do not need to be at home to take tests or answer questions; one family member answers basic questions (such as working status and school enrollment) for the entire household. In addition, in the long-term analysis using the NSS data, people who experienced higher rainfall at particular ages have lower overall schooling, which is consistent with the dropout rates we observe in the ASER sample.

2.5.5 Selective Migration in Early Life Regressions

The sort of selective migration that could bias our early life regressions in Table 2.5 is somewhat different. Even if migration patterns are driven by rainfall patterns, as long as these migration patterns are not age specific, then they would not bias our estimated coefficients. In the context of our early life results, this is reasonable. For instance, even if children exposed to drought conditions under the age of two are more likely to move (and those who move are positively selected biasing our results upward), they would likely move with their whole family including older and younger siblings. Thus, each "treatment" child would likely travel with several "control" children. In our main specification in Table 2.5, we use household fixed effects which means that the child is only compared to the other children in his household mitigating any concerns that household migration could be driving our results.

In the long term results in Table 2.7, our main finding is that rainfall shocks around the ages of 11-15 matter for later life outcomes. In the NSS and the ASER data, we assume that the district in which an individual is surveyed is the district in which he spent those years. As stated above, cross district migration is not terribly common in India, and to the extent that it is orthogonal to drought exposure in childhood, it will simply attenuate our results. However, if children are systematically moving out of districts in which there is low rainfall when they are leaving school, this could bias our results. However, again to the extent that these migrants are positively selected this will bias our results downward, since high rainfall at puberty is negatively associated with later life outcomes.

It is also important to remember that rainfall shocks are defined as the top and bottom quintile of rainfall, respectively. The average child will experience 2 or 3 “droughts” by this definition over the course of his childhood, and it is unlikely that he is leaving the district in response to relatively small productivity fluctuations.

2.5.6 Selective Fertility and Mortality

In the early life analysis, one potential concern with trying to understand the effect of drought on cognitive development is that we only observe children who survive and make it into the sample; if drought exposure increases infant and early childhood mortality, it could affect the composition of our sample in “control” and “treatment” years. This selection would most likely bias our results downward; since these are the children who survived, they are positively selected and probably do better on health and educational outcomes relative to the children who died off. Therefore, we are less concerned about bias from selective mortality.

However, another potential concern with the early life results could be if women are delaying and/or changing fertility patterns in response to droughts. For example, mothers may choose to wait out a drought year before having a child. If droughts are in fact impacting fertility decisions, the empirical results could be biased upward if the children being born in drought years are negatively selected.

Since our dataset includes only children ages 5-16, both of these selection effects would show up as smaller cohort sizes observed for treatment cohorts (assuming that most of the selective mortality happens before age 3). Unfortunately, population by district is only available every 10 years from census data. Therefore we investigate the issue of selective fertility for individuals born in 1991 or 2001 (since that is when census data is available). We regress the ln number of children in each cohort by district on measures of drought and ln population by district in Table B.20. In column 3 instead of total population, we use female population ages 15-49 from the 2001 Census since this is the relevant childbearing population. Given we are not exactly sure when mothers and fathers make decisions about when to conceive, we investigate the period 5 years prior to birth.

Table B.20 reports the results of these OLS regressions for 1991 and 2001. Most of the coefficients are small, and none are statistically significant in 1991. In columns 2-3, drought in *t-4* is significantly (and positively) correlated with number of births. However, none of the other coefficients are statistically significant. These data do not suggest that there is a systematic difference in the size of "treated" cohorts, and thus selective fertility and mortality are unlikely to be driving our results. Recall also that these are not necessarily severe droughts in that they are defined as rainfall below 20th percentile within the district.

Another piece of evidence which points against selective fertility (and selective migration) are the household fixed effects results of Table 2.5. If either of these mechanisms is driving the results, then within household variation in drought exposure should not affect cognitive test scores. This story relies on *between* household variation—i.e. that "good" households are acting differently with respect to droughts compared to "bad" households. That is, if "good households" are leaving the area after droughts, or delaying their fertility when there are droughts, then our sample of exposed children would be more heavily weighted toward "bad households" which could bias our results upward. However, the results with and without household fixed effects are extremely similar (results without household fixed effects that include district fixed effects are available upon request), which leads us to conclude that this type of selection is unlikely to be biasing the estimates.

2.6 Discussion and Conclusion

In this paper we present a simple model of human capital investment, and show that under reasonable conditions, we would expect the effect of wages on human capital investment to be negative when children are school-aged. We also show that, in the presence of strong dynamic complementarities, early life wages will positively affect investment and schooling and overall human capital.

We estimate these comparative statics using test scores, schooling outcomes, and labor market data from rural India. We show that positive productivity shocks cause lower school enrollment and attendance, and lower overall test scores. We argue that this is due to children substituting from human capital producing activities to outside work or home production when wages are high, using evidence from the NSS labor market survey on children's reported activities.

In addition, we show that the lagged effects of early life positive rainfall shocks on both schooling and human capital are positive. Children who were exposed to droughts in early life score significantly worse on math and reading tests, and are more likely to be behind in school or to never have enrolled. According to our model, this is evidence of dynamic complementarities in the human capital production function: the early life investments in these children (due to increased consumption) increase not just the level of human capital but also the return to additional human capital investments.

It is important to note that our model assumes that schooling has no direct costs, and that there is sufficient scope for substitution from schoolwork to productive work either in the home or in the labor market. In particular, school fees together with liquidity constraints could cause substitution away from schooling during lower wage years even if the assumptions of our strictest model hold. These assumptions are reasonable in India, but may differ in other developing country settings. Our findings are consistent with a growing literature about the effect of wages on time-intensive investments in children more generally (Atkin (2012), Miller and Urdinola (2010)).

These results indicate that opportunity costs of human capital investment matter even for young children, and that higher wages for low education jobs could have the counterintuitive effect of lowering human capital investments in children. This research could inform policy decisions about poverty alleviation programs. Many poverty alleviation programs in the developing world take the form of work programs with inflated wages for agricultural laborers. For example, NREGA in India generated 2.57 billion person days of employment (in 2010-11). If these types of programs raise prevailing wages, they could cause students to substitute toward work and away from for school attendance, even if the programs are only in place for adults. In fact, Shah and Steinberg (2014) show that NREGA increases rural wages and decreases human capital investment, especially for older children. Lump sum grants or even conditional cash transfers might be better options in this context.

Though these results focus on productivity fluctuations rather than steady growth, they indicate that the reaction to wage growth in low income areas could be to *decrease* investment in human capital which could be detrimental to long term growth and poverty reduction. If poor countries want to increase school enrollment and attendance, they should not only consider fees and tuition, but the opportunity cost of attendance in terms of wages as well.

Chapter 3

Workfare and Human Capital: Evidence from NREGA in India[1]

3.1 Introduction

Workfare programs are an increasingly popular option for alleviating poverty in developing countries. These programs typically provide aid to the poor with corresponding work at a guaranteed wage. They have become a popular means of redistribution as they provide aid to the poor along with corresponding work incentives. Recent examples include programs in Malawi, Bangladesh, India, Philippines, Zambia, Ethiopia, Sri Lanka, Chile, Uganda, and Tanzania. In fact, the practice of imposing work requirements for welfare programs stretches back at least to the British Poor Law of 1834 (Imbert and Papp, 2012b).

It is important to note that government-provided workfare programs affect not only overall income but also the prevailing wage and time cost of family members. Ravallion (1987) and Basu *et al.* (2009) show that government hiring may crowd out private sector work and lead to a rise in equilibrium private sector wages. It is possible such programs could lead to decreased human capital production for certain individuals. Shah and Steinberg (2014) show that the opportunity cost of human capital investment matters even for young

[1] Co-authored with Manisha Shah

children, and that higher wages for low education jobs could have the counterintuitive effect of lowering human capital investments in children. They find that when labor market opportunities and wages improve in rural India, children (ages 11-15) are more likely to drop out of school and participate in labor markets. This is similar to findings by Atkin (2012), who finds that factory openings in rural Mexico increase dropout rates, and Miller and Urdinola (2010) who finds that increased wages in rural Colombia decrease time-intensive health investments in children.

In this paper we study the impacts of NREGA, one of the largest workfare programs in the world, on human capital outcomes of children ages 5-16. We exploit the three-phase rollout of NREGA, a program introduced gradually throughout India starting with the poorest districts in early 2006 and extending to the entire country by 2008. We find that children score significantly lower on math and reading test scores once NREGA enters their district, and they are also significantly more likely to drop out and less likely to both attend and be on track in school. These results are primarily driven by children ages 13-17 which is precisely the age group that is most likely to enter the labor market. We argue that these results are caused by increases in labor demand, which increase the opportunity cost of schooling for children. We find that though the results on human capital are similar for both boys and girls, girls are more likely to substitute for their mothers in domestic work, while boys are more likely to work outside the home for pay.

This paper contributes to the literature in three ways. First, as far as we know, this is the first paper to document the possibility that a workfare program can lead to lower levels of human capital attainment directly using nationally representative test scores and school enrollment rates.[2] Secondly, this research adds to the growing literature about the unintended price impacts of social programs more generally (including Jayachandran *et al.* (2013) and Hastings and Washington (2010), among many others). Third, these findings could be of direct interest to policy makers considering using workfare as a means of poverty

[2]Li and Sehkri (2013) examines the impact of NREGA on school enrollment numbers using the DISE data. Their findings are broadly consistent with our own.

alleviation. If these types of programs raise prevailing wages and cause older students to substitute toward work and away from school, lump sum grants or conditional cash transfers might be better options in this context.

This paper is organized as follows. Section 2 describes the NREGA program and the various datasets we use in this paper. Section 3 provides the empirical framework and Section 4 the results and robustness. Section 5 presents mechanisms and we conclude in Section 6.

3.2 Background and Data

3.2.1 Background on NREGA

The National Rural Employment Guarantee Act (2005) of India provides a legal guarantee of up to 100 days of annual employment at the statutory minimum wage rate[3] to rural households willing to supply manual labor on local public works in a financial year (Ministry of Rural Development, 2005). The Act mandates equality of wages for men and women and one-third of program beneficiaries to be women. It is operationalized through the National Rural Employment Guarantee Scheme (NREGS) which began in 2006 and has an annual budget of around Rs. 48,000 crores (approx. 9 billion dollars), amounting to more than 11% of the 2011 Union budget expenditure.

To obtain work on a project, interested adult members of a rural household must apply for a Job Card at the local Gram Panchayat.[4] After due verification, the Gram Panchayat issues a Job Card, and the card should be issued within 15 days of application. The Job Card bears the photograph of all adult members of the household willing to work under NREGA and is free of cost. Workers can apply for work at any time once they have a job card. The applicants must be assigned to a project within 15 days of submitting the application. If

[3]The statutory minimum wage rate varies across states but it is approximately 2USD per day.

[4]The Gram Panchayat is the lowest level of administration in the Indian government comprising a group of villages.

they are not given a job, they are eligible for unemployment compensation. Applicants have no choice over the project. The particular types of projects allowed under NREGA are typical rural employment projects such as road construction, earthworks related to irrigation and agriculture, and water conservation. The federal government bears the entire cost of wages of the workers and 75 percent of the cost of materials. The state governments bear the remaining 25 percent. In addition, state governments bear the cost of unemployment allowance payable when the state government cannot provide wage employment on time (Azam, 2012; Ministry of Rural Development, 2008).

3.2.2 Rollout of NREGS

The National Rural Employment Guarantee Act was passed in 2005, and the scheme begun rollout in February 2006. At this time, 200 districts were given access to the program. In April 2007, a further 130 districts were added, and in April 2008, the program became available in the remaining 270 districts. In this paper, we will refer to these groups of districts as “wave 1,” “wave 2,” and “wave 3,” respectively. While the actual assignment mechanism to waves is unknown, the government stated an explicit goal of rolling out the program to the poorest districts first. However, it was also guaranteed that each state would have at least one district in the first wave of the program. Zimmerman (2014) argues that based on the allocation of similar programs, it was likely that the states were given slots to allocate to each wave based on their levels of poverty, and that their allocation was likely based on the government’s own “backwardness rankings" (based on agricultural wages, % SC/ST and agricultural productivity).

Given that the rollout was explicitly designed to be non-random (needier districts received the program first), we cannot simply compare outcomes across waves. Instead, we will rely on the assumption that the program was rolled out based on static characteristics of the districts, rather than underlying trends in school enrollment or child labor. We will test this assumption explicitly in Section 3.4.2.

3.2.3 Cognitive Testing and Schooling Data

Every year since 2005, the NGO Pratham has implemented the Annual Status of Education Report (ASER), a survey on educational achievement of primary school children in India which reaches every rural district in the country.[5] We have data on children for 2005-2009, giving us a sample size of approximately 2.5 million rural children. Thus, we have one round of pre-NREGA test score data (2005), three rounds of test score data during the rollout (2006-2008), and one round after the program has been rolled out throughout the country (2009). The sample is a representative repeated cross section at the district level. The ASER data is unique in that its sample is extremely large and includes both in and out of school children. Since cognitive tests are usually administered in schools, data on test scores is necessarily limited to the sample of children who are enrolled in school (and present when the test is given). However, ASER includes children ages 5-16, who are currently enrolled, dropped out, or have never enrolled in school. In fact, Pratham's explicit goal is to test children who are both in and out of school. They survey at the household on Sundays, when people generally do not work and children are not in school, and return to households where children are not present at the time of the survey. In Table 3.1 we describe the characteristics of the children in our sample as well as their test scores.

The ASER surveyors ask each child four to six math questions. The three math questions asked every year are whether the child can recognize numbers, subtract, and divide. The scores are coded as 1 if the child correctly answers the question, and 0 otherwise. The first math score variable varies from 0 to 3 depending on how many tasks the child can complete correctly. In 2006 and 2007, children were also asked two word problems. In the second math score measure, we combine the first math score and the math word problems. This is a variable equal to the total number of correct answers given by the child and ranges from 0-5.

Each child was also asked to complete four reading tasks. The child is tested as to

[5]This includes over 570 districts, 15,000 villages, 300,000 households and 700,000 children in a given year. ASER is the largest annual data collection effort with children in India. For more information on ASER, see `http://www.asercentre.org/ngo-education-india.php?p=ASER+survey`

Table 3.1: *Summary Statistics*

ASER Summary Statistics (Ages 5-16)			
	Mean	Std. Dev.	Observations
Age	10.2	3.2	2,666,336
Female	.45	.50	2,666,336
Math Score	1.80	1.05	2,666,336
Math Score with Word Problems	2.21	1.54	2,666,336
Reading Score	2.71	1.55	2,681,363
Currently Enrolled	.94	.25	2,768,347
Current Grade	4.6	2,581,262	
ASER Summary Statistics (Ages 13-16)			
	Mean	Std. Dev.	Observations
Age	14.3	1.08	720,903
Female	.45	.50	720,903
Math Score	2.48	.846	720,903
Math Score with Word Problems	3.17	1.43	720,903
Reading Score	3.57	.973	723,264
Currently Enrolled	.87	.34	765,647
Current Grade	7.66	2.53	677,122
NSS Sample (Ages 13-17)			
	Mean	Std. Dev.	Observations
Age	14.97	1.36	183,218
Female	.47	.49	183,218
Primary Activity:			
Attends School	.68	.47	183,218
Works at Home	.09	.27	183,218
Works outside Home	.07	.26	183,218
Domestic Work	.11	.31	183,218
Young Lives Sample (Older Cohort, Ages 11-15)			
	Mean	Std. Dev.	Observations
Share of Time Spent in School	.25	.13	748
Share of Time Spent on HH Chores	.05	.05	748

Notes: This table shows summary statistics from the ASER, NSS, and Young Lives data.

whether he/she can recognize letters, recognize words, read a paragraph, and read a story. We generate a total reading score that varies from 0 to 4 depending on how many tasks the child can complete. The ASER dataset also includes information on whether the child is currently enrolled in school and the current grade of the child which we use as additional measure of schooling.[6]

3.2.4 NSS Data

To examine the impact of NREGA on work and wages, we use the NSS (National Sample Survey) Round 60, 61, 62, 64 and 66 of the NSS data which was collected between 2004 and 2009 by the Government of India's Ministry of Statistics. This is a national labor and employment survey collected at the household level all over India, and we use data from all rural households in these surveys. Rounds 60 and 61 (2003 and 2004) are pre-rollout, round 62 straddles 2005 and 2006 but very few districts are sampled in 2006, round 64 is collected during the rollout (2007), and round 66 is collected after the rollout (2009)[7].

This dataset gives us measures of employment and schooling status at the individual level. The survey asks what the "primary activity" of each member of the household is. We define "domestic work" as individuals who report to attending domestic duties and/or engaging in free collection of goods (vegetables, roots, firewood, cattle feed, etc.), sewing, tailoring, weaving, etc. for household use; we define "works at home" as individuals who self-report to being self-employed and working in their household enterprise (either as an own account worker, an employer or as an unpaid family worker). We define "works outside of home" as someone who reports their status as a regular salaried/ wage employee or as a paid casual laborer.

[6]More information on the ASER survey questions, sampling, and procedures can be found in the ASER data appendix.

[7]For the placebo analysis and migration analysis we also use round 55 (1999-2000) of the NSS data. This is necessary as it provides a longer time horizon for pre-trends, and is one of the only rounds (other than 2007) to ask about temporary migration.

3.2.5 Young Lives Data

The Young Lives data is a household level panel dataset from Andhra Pradesh which surveyed two cohorts of children in 2002 (Round 1), 2006 (Round 2), and 2009 (Round 3). It is following the lives of children in two age-groups: a Younger Cohort of 2,000 children who were aged between 6 and 18 months when Round 1 of the survey was carried out in 2002, and an Older Cohort of 1,000 children then aged between 7.5 and 8.5 years. The survey was carried out again in late 2006 and in 2009. The children were sampled in geographic clusters, which were selected through a semi-purposive approach. Within each cluster, children were randomly selected.

Young Lives collected detailed time use data for both cohorts of children in Rounds 2 and 3 which we use in this paper. The younger cohort was 4-5 in Round 2 and the older cohort was 11-12 years old. In Round 3, the younger cohort was 7-8 and the older cohort was 14-15 years old. We restrict our sample to rural children which is about 75 percent of the sample.

3.3 Empirical Strategy

Because NREGA was rolled out in three waves, we can compare districts before and after the program was in place while controlling for overall year and district effects. This strategy allows us to identify effects off the quasi-random timing of the rollout, rather than which districts got NREGA first, which of course was targeted to alleviate poverty.

We estimate the following regression:

$$S_{ijty} = \alpha + \beta_1 \delta_{j,t} + \gamma_j + \phi_t + \psi_y + \epsilon_{ijty} \tag{3.1}$$

where S_{ijty} is a schooling outcome variable (such as test scores or enrollment status) for child i in district j in year t who is age y, $\delta_{j,y}$ is an indicator of whether district j had NREGA in year t, γ_j is a vector of district fixed effects, ϕ_t is a vector of year fixed effects, and ψ_y is a vector of child age fixed effects. β_1 is our coefficient of interest, and it measures the effect of NREGA on the outcome variable S. Our standard errors are clustered at the district-year level.

3.4 Results

3.4.1 Main Results

Table 3.2 shows our primary estimates of β from equation 1 in the ASER data. Panel A shows results for the full sample of children in the ASER data, age 5-16. Columns 1-3 show the effect of NREGA on test scores: while all three coefficients are negative, the coefficient on the first math score and the reading score are small and not significant. The coefficient on the second math score is -.1, and statistically significant at the 1% level. This represents about a 5% increase from a mean of 1.84. Column 4 shows our estimate of the effect of NREGA on the probability that a child reports being enrolled in school. This coefficient is negative and statistically significant, and indicates that NREGA increased the probability that a child was out of school by .6 percentage points. Finally, column 5 shows our estimates of the effect of NREGS on a child's highest reported grade, which was reduced by .09 years (or 2%). Broadly, these results are consistent with NREGS reducing overall human capital investment, both on the intensive and extensive margins.

In Panels B and C of Table 3.2, we separate children into two age groups: primary school age (5-12), and secondary school age (13-16). All of the coefficients in Panel B (older children) are larger than those in Panel C, though they are not always statistically distinguishable from one another. Enrollment rates for older children are dropping by almost a full percentage point when NREGA rolls out, which represents an 8% increase in the probability that these children are out of school. Current grade also decreases by .1 year when NREGA rolls our for this older group. In addition, all three test scores have negative and statistically significant coefficients, which indicates that learning, not just reported enrollment rates, are going down for these children as well. The results in Panel C are much smaller and generally not statistically significant. Therefore, it seems the declines in observed human

Table 3.2: *Effect of NREGA on Test Scores and Schooling*

Dep. Var:	Math Score	Math Score2	Reading Score	Currently Enrolled	Current Grade
			Panel A: Full Sample		
NREGA	-.01	-.1	-.001	-.006	-.09
	(.01)	(.02)***	(.02)	(.002)**	(.02)***
Observations	2,666,336	2,666,336	2,681,363	2,768,347	2,581,262
Mean DV	1.80	2.21	2.71	.94	4.60
			Panel B: Ages 13-16		
NREGA	-.02	-.13	-.03	-.009	-.1
	(.01)*	(.02)***	(.01)**	(.004)**	(.03)***
Observations	720,903	720,903	723,264	765,647	677,122
Mean DV	2.48	3.17	3.57	.87	7.66
			Panel C: Ages 5-12		
NREGA	.0008	-.05	.02	-.005	-.02
	(.01)	(.02)***	(.02)	(.002)**	(.02)
Observations	1,945,433	1,945,433	1,958,099	2,002,700	1,904,140
Mean DV	1.55	1.84	2.40	.96	3.51
Dist FEs	YES	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on math and reading test scores and schooling outcomes from the ASER data. All columns contain data from 2005-2009. Math score ranges from 0-3 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide(=3), Math Score2 ranges from 0-5 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide (=3), Can do first word problem (=4), Can do second word problem 2 (=5), Read score ranges from 0-4 (Nothing(=0), Read Letters(=1), Read Words(=2), Read Paragraph(=3) and Read Story(=4). All Panels include year and district fixed effects. Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table 3.3: *Effect of NREGA on Test Scores and Schooling By Gender*

Dep. Var:	Math Score	Math Score2	Reading Score	Currently Enrolled	Current Grade
			Panel A: Females, Ages 13-16		
NREGA	-.02	-.15	-.04	-.008	-.1
	(.02)	(.03)***	(.02)**	(.004)*	(.04)***
Observations	323,846	323,846	325,018	343,696	302,034
Mean DV	2.42	3.09	3.53	.86	7.60
			Panel B: Males, Ages 13-16		
NREGA	-.02	-.11	-.03	-.008	-.1
	(.01)	(.02)***	(.01)**	(.004)**	(.04)***
Observations	397,057	397,057	398,246	421,951	375,088
Mean DV	2.53	3.24	3.60	.88	7.71
Dist FEs	YES	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on math and reading test scores and schooling outcomes from the ASER data. All columns contain data from 2005-2009. Math score ranges from 0-3 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide(=3), Math Score2 ranges from 0-5 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide (=3), Can do first word problem (=4), Can do second word problem 2 (=5), Read score ranges from 0-4 (Nothing(=0), Read Letters(=1), Read Words(=2), Read Paragraph(=3) and Read Story(=4). All Panels include year and district fixed effects. Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

capital are being driven by the older children.

In Table 3.3, we break down the results for older children by gender. The coefficients for boys and girls are very similar across all outcome variables. Though the coefficients on two of the three test scores (math2 and reading) are slightly larger for girls, the estimates are not statistically distinguishable from one another. Overall, it seems from these results that NREGS is affecting human capital investment equally for adolescent girls and boys.

While these results are informative about the reduced form effect of the program on human capital investment, we cannot learn whether children are substituting into productive work (as hypothesized) with the ASER data, since we have no information on their other activities. In Table 3.4, we show our estimates of the effect of NREGS on children's reported "primary activity" using the NSS data. Here, we restrict to children age 13-17, though we report coefficients for younger children in the appendix. Columns 1 and 2 show the estimates of NREGS on whether the child reports that his or her primary activity is productive work

Table 3.4: *Effect of NREGA on Working and School Attendance (NSS)*

	Ages 13-17			
Years:	2003-2009 (Actual)		1999-2005 (Placebo)	
Dep. Var:	Works	Attends School	Works	Attends School
NREGA	.03	-.03	-.005	-.005
	(.008)***	(.008)***	(.009)	(.009)
Observations	183,218	183,218	115,553	115,553
Mean of DV	.25	.68	.29	.66
Dist FEs	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on children working and attending school. Columns 1-2 use data from the NSS 2003-2009 (rounds 60, 61, 62, 64, 66). The placebo analysis utilizes NSS 1999-2005 data (rounds 55, 60, 61, 62) where NREGA equals 1 in 2004. Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

versus attending school. Here, we find that children are three percentage points less likely to report attending school, and equally more likely to report working.[8]

3.4.2 Robustness

Our identification relies on the assumption that in the absence of NREGA, the districts that received the program earlier and those that received the program later did not have systematically different time patterns in our outcome variables.[9] While we cannot test this assumption directly, we can do two robustness tests to see whether it appears as though our results are being driven by the program or by underlying trends in eligible districts.

In the NSS data, we can analyze the trends in outcomes leading up to the passage of NREGA for early versus late access districts. In columns 3 and 4 of Table 3.4, we report results from a placebo regression in which we falsely assign early districts (waves 1 and 2) to have NREGA during 2004, and late districts (wave 3) to get the program in 2005 (in reality,

[8]Part of the discrepancy in magnitudes between these results and the ASER results might be due to how the question is asked. In the ASER, parents are simply asked whether their child is enrolled in school, and if so, which type. In the NSS, parents are asked what the primary activity of the child is.

[9]This is similar to the notion of "parallel trends" in differences-in-differences, except that since we have observations from more than two points in time, the timing effects need not be linear. In other words, conditional on district and year fixed effects, district-year observations that had access to NREGA differed only due to the program, not due to differential effects of year by wave.

no district got the program until 2006).[10] Reassuringly, we find no effect of the program on primary activity before it began, and we can reject that the magnitudes are equal to our main effect size at the 5% level. This placebo treatment does not predict differences in probability of working or attending school.

For our analysis of the ASER data we cannot examine pre-trends because our data begins in 2005, one year before the program was rolled out in wave 1 districts. However, since we have 5 years of data and 3 separate roll-out groups, we can estimate our main analysis including wave-specific linear time trends. If our results are picking up gradual differences in outcomes over the 5-year period, then the waves trends should absorb this effect, and our coefficients should go to zero. However, if the impact is occurring due to the change from NREGA, the coefficients should remain the same. In Table C.1, we show estimates of our coefficients from Table 3.2 with the addition of wave-specific linear time trends. Reassuringly, the coefficients are all still negative, and similar in magnitude, though some are less precisely estimated.

In addition, there was substantial heterogeneity in the size and scope of the program when it was introduced. Some districts were particularly good at implementation, and Imbert and Papp (2012a) argues that these differences did not reflect differences in the underlying supply of labor. It seems plausible that districts which had a larger number of NREGA jobs would be more likely to experience decreases in human capital. To examine this, we create a dummy variable which is equal to one if the district is above the median number of jobs created in 2009, relative to population in the 2011 census. We choose 2009 because it is the earliest year in which all districts had NREGA for at least a year.[11] In Table C.3, we regress the ASER human capital outcomes on this variable interacted with NREGA availability. Though the results of these regressions are not precise, the interaction terms are

[10] We combine waves 1 and 2 in this analysis because in our main NSS results, we do not have much data from 2006. Thus, our identification comes from the differences between the districts in waves 1 and 2 and the districts in wave 3. The results are robust to assigning treatment to districts in wave 1 in 2004 and wave 2 in 2005.

[11] Total job cards supplied is highly correlated across years.

consistently negative, and often larger than the NREGA dummy, which indicates that the results are primarily being driven by districts which had more NREGA jobs. The lack of precision in these regressions, as well as the endogenous nature of the interaction variable, leads us to take these results as suggestive, rather than conclusive.

3.5 Mechanisms

In the previous sections, we have shown that the introduction of NREGA caused a decrease in human capital investment amongst adolescents. In this section, we will outline the evidence for several possible mechanisms. First, we will examine whether NREGA created a shift in labor demand that increased the opportunity cost of schooling. In addition, we will assess whether changes in the returns to schooling, decreased parental supervision, or selective migration could be driving our results.

3.5.1 Labor Demand

From previous work, we know that NREGA caused an increase in labor demand in the districts in which it was operational (Imbert and Papp, 2012a; Zimmerman, 2014; Azam, 2012). Though work on NREGA projects was legally limited to those over the age of 18, this could have caused an increase in labor demand for adolescents in a few ways. First, there could have been some leakage in who was allowed to work for the program, with either adolescents lying about their age, or program administrators looking the other way. Second, the introduction of NREGA jobs could create additional jobs, such as selling tea or food to workers. Lastly, adolescents' labor could be substitutes for adults' labor, so that when adults begin working for NREGA, adolescents take their places doing household, farm, and domestic work.

While we know from Table 3.4 that children reported increases in productive work at the onset of NREGS, in order to understand what is causing this, we look more closely at the shifts in labor amongst both parents and adults. In Table 3.5, we show the effect

Table 3.5: *Effect of NREGA on Parents' Primary Activities*

Panel A: Mothers			
Dep. Var:	Works at Home	Works outside Home	Domestic Work
NREGA	.02	.024	-.039
	(.008)**	(.006)***	(.01)***
Observations	228,747	228,747	228,747
Mean of DV	.22	.14	.62
Panel B: Fathers			
Dep. Var:	Works at Home	Works outside Home	Domestic Work
NREGA	-.035	.051	-.003
	(.008)***	(.008)***	(.003)**
Observations	235,086	235,086	235,086
Mean of DV	.54	.41	.02
Dist FEs	YES	YES	YES
Year FEs	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA . Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

of NREGA on the reported "primary activity" for mothers and fathers using NSS data.[12] Column 1 shows the effect of NREGS on working in a home enterprise, either as its head, an employer, or an unpaid family worker.[13] We find that for mothers, this work is increasing by 2 percentage points, while for fathers, it decreases by 3.5 percentage points. Working outside the home is increasing for both mothers and fathers, though the increase is larger for fathers (5 percentage points, versus 2.4 percentage points). Lastly, domestic work is decreasing for both spouses, though the magnitude is only really meaningful for mothers (3.9 percentage points). Overall this is consistent with mothers switching out of domestic work and into market work, either at home on the farm or outside in the market; while fathers are switching from working at home to working for wages in the market.[14]

[12]Male heads of household age 18-64 and female spouses of household heads age 18-64, respectively.

[13]In our setting, these household enterprises will primarily be farms.

[14]NSS does not directly ask households in 2007 if they are working for NREGA or not, so we do not know whether the switch to outside work is into NREGA jobs or other paid work.

Table 3.6: *Children's Primary Activity (Ages 13-17)*

Panel A: All Children				
Dep. Var:	Works at Home	Works outside Home	Domestic Work	Attends School
NREGA	-.005	.02	.01	-.03
	(.004)	(.004)***	(.005)***	(.008)***
Observations	183,218	183,218	183,218	183,218
Mean of DV	.09	.07	.11	.68
Panel B: Girls				
NREGA	.0008	.003	.03	-.03
	(.005)	(.005)	(.009)***	(.01)**
Observations	85,742	85,742	85,742	85,742
Mean of DV	.06	.05	.22	.63
Panel C: Boys				
NREGA	-.01	.03	.006	-.03
	(.006)*	(.006)***	(.003)**	(.009)***
Observations	97,476	97,476	97,476	97,476
Mean of DV	.11	.10	.01	.72
Dist FEs	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA . Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

In Table 3.6, we show our estimates of the impact of NREGA on primary activity, broken down by the same categories for children ages 13-17.[15] Panel A reports the results for all children, while Panels B and C report the effects for girls and boys, respectively. While we do not observe differences in the effects of NREGA on human capital investments by gender (both boys and girls are three percentage points less likely to report attending school), we do observe differences in work type by gender. The pattern for boys is similar to that of adult men: they reduce working at home (and attending school), and increase the probability of working outside the home. Thus, for boys it looks as though they are either working directly for NREGA, or in market jobs as substitutes for adult labor.

For girls, the results are quite different. Girls appear to be substituting almost entirely

[15]In Appendix Table C.4, we report these results for children age 5-12. We do not see similar effects for them. Indeed, if anything, their school attendance might be increasing slightly.

into domestic work when leaving school. Since mothers are substantially decreasing domestic work when NREGS becomes available, it is likely that these girls are substituting for their mothers' labor inside the household. While NREGS was intended to be empowering for women by encouraging their participation in market work and offering equal wages, it might also be keeping adolescent girls from either school or market work, and increasing the probability they spend time cooking and cleaning at home. We show these results graphically in Figures 3.1 and 3.2, respectively.

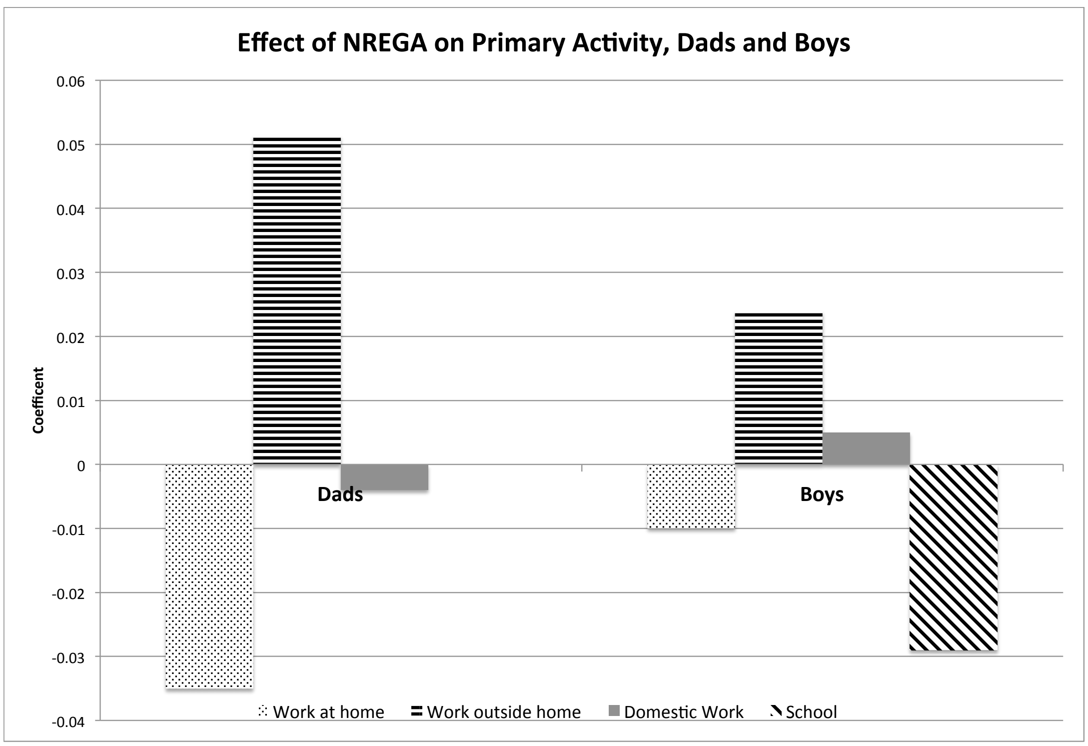


Figure 3.1: *Effect of NREGA on Primary Activity, Dads and Boys*

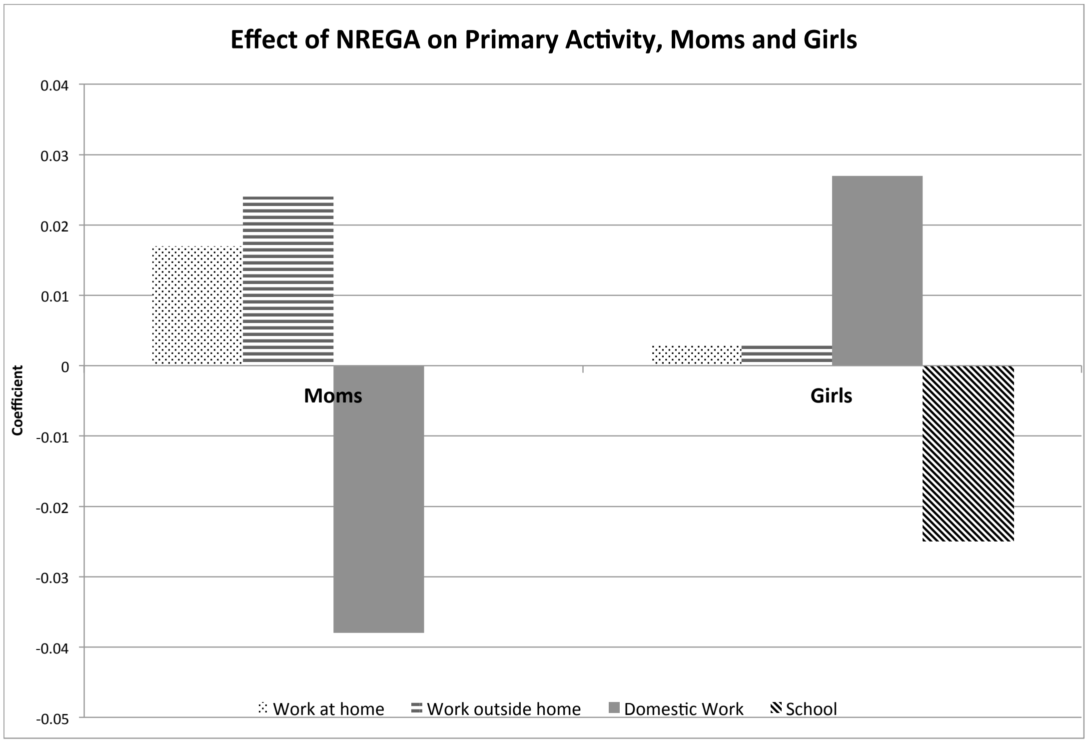


Figure 3.2: *Effect of NREGA on Primary Activity, Moms and Girls*

To supplement this analysis, we turn to the Young Lives data from Andhra Pradesh. While this data is only from one state and includes only two cohorts of children, it asks detailed information on time use. The time use data is available in rounds 2 and 3, and round 2 took place in 2006, when wave 1 districts had access to NREGA, and waves 2 and 3 districts did not and round 3 took place when all districts had been phased in. Since this is a panel of two cohorts of children, we can estimate child fixed effects models.

In Panels A of Table 3.7, we show the effect of NREGS on time spent in school and time spent on household chores for all the children, and in panels B and C we split the sample by older and younger cohorts. Consistent with our earlier findings, NREGS decreases time

Table 3.7: *Effect of NREGA on Schooling and Time Use*

Dep. Var:	School Time	HH Chore Time
Panel A: Full Sample		
NREGA	-.02	.01
	(.008)***	(.003)***
Observations	2205	2205
Mean DV	0.27	0.02
Panel B: Older Cohort		
NREGA	-.03	.02
	(.02)**	(.007)***
Observations	748	748
Mean DV	.25	.05
Panel C: Younger Cohort		
NREGA	-.02	.001
	(.008)*	(.002)
Observations	1457	1457
Mean DV	.28	.009
Child FEs	YES	YES
Round FEs	YES	YES

Notes: This table shows estimates of the effect of NREGA on schooling outcomes and time use from the Young Lives data. This is a panel of 2 cohorts of children (young and old): round 1 is from 2002, round 2 from 2006 and round 3 from 2009. The time use questions were asked in rounds 2 and 3 for both cohorts. They are hours/day in a 24 hour period spent in each activity. All regressions include child and round fixed effects. Standard errors clustered at the child level and are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

spent in school by 2-3% and increases time spent on household chores by 1-2%. These results are similar for both genders and much stronger for the older cohort. In Panel C, we show these estimates for the younger cohort, who are about 5 in 2006. The coefficients on school time and household chores are in the same direction as for the older children, but they are smaller and household chores is not statistically significant.

3.5.2 Alternative Explanations

Returns to Schooling

Since NREGA is a transfer program to the poor and largely uneducated, it can reduce the returns to schooling. If families are forward looking, they could reduce their schooling investment to adjust to the new, lower returns, which could explain our findings. While this is possible, we think it is unlikely for a few reasons. First, the program was passed and announced in 2005, with a lot of press. If families really are forward-looking enough to adjust their human capital investment to a decrease in the return to schooling, they should adjust fully upon the announcement of the program, not when it is rolled out to their district. Our identification relies on the differential impact of NREGA when it is rolled out, so for returns to schooling to be the driving cause of the lowered human capital investment, it would need to be the case that families updated fully only when the program actually came to them, not when they found out about it.

In addition, we think it is unlikely that this mechanism is driving out results because the changes in agricultural incomes are simply too small. The increase in the agricultural wages due to NREGS is about 5 percent (Imbert and Papp (2012a); Berg *et al.* (2012); Azam (2012)). To compare this with the differences in incomes by education, the average wage for those with less than 8 years of schooling is 100 rupees (1.50 USD) per day (2003-2008). According to the literature, NREGS would have increased this to 105. For those with at least 12 years of schooling, the average wages is approximately 350 rupees per day. While this is not necessarily the causal return to secondary school for the marginal student, the magnitude swamps that of any changes in income due to this program. Adding this to the uncertainty that the program will continue in its present incarnation long enough for these students to reap its benefits, makes it seem unlikely that the changes in returns to schooling would be driving our results.

Parental Supervision

Another possible channel through which NREGA could impact human capital investment is if parents are integral in ensuring that their kids show up to school. If many children go from having mothers who work primarily in the home to mothers working outside the home, the lack of supervision could allow them to stay home from school without detection from their parents (or, perhaps, there is no one there to get the child to and from school). While it is certainly possible that some of this is going on, we think it is unlikely that this effect is driving our results. First, we primarily see the reduction in schooling for older children, age 13-17. These children are almost surely not being walked to school by their mothers and could probably skip school even with mothers at home. Second, we see commensurate increases in domestic work and work outside the home on surveys administered to the head of household. That is, parents are reporting to us that their children are not in school, but rather engaging in productive work. This seems incongruous with the idea that children are simply sneaking around when their parents are out of the house working.

Selective Migration

Lastly, there is evidence that NREGA might decrease short-term migration from rural to urban areas for work (Imbert and Papp, 2014). This could bias the sample of children observed when enumerators survey the village and affect our analysis of test scores and school enrollment. However, it is important to note that to the extent that migrants are positively selected, we would expect this to bias our results downward, since these children would be more likely to show up in our sample.

In Table C.6, we test whether the adolescents in our sample are less likely to migrate out for temporary work when NREGS is rolled out. If anything it looks as if the opposite is true, though both overall migration rates and differences due to NREGS are very small. For this age group, temporary migration is increasing by .9 percentage points due to NREGA, from an average of 1.8 percent. This is entirely driven by boys.

In addition, temporary migration for work is limited almost entirely to males in India. Women tend to stay in their parents' home until marriage, when they move to the home of their husband's family. Thus, if migration were driving our results (presumably through negative selection of migrants, which again, seems unlikely), we should expect to see this effect only in boys. However, in our results, adolescent girls experience similar reduction in test scores, enrollment rates, and grade, as well as commensurate increases in productive work. For this reason, we think it is unlikely that our results are an artifact of NREGA-induced selective migration.

3.6 Discussion and Conclusion

This paper shows examines the effect of NREGA, a large workfare program in India, on school enrollment, child labor, and test scores. We show that NREGA decreased human capital investment, primarily for children over the age of twelve, and that this was likely caused by boys responding to the increase in labor demand by working outside the home, and girls substituting for their mothers in domestic work. These results are extremely consistent with earlier findings on the effect of wages on human capital investment in India (Shah and Steinberg, 2014), though these results might be of more interest to policy makers, since the wage increase is being directly caused by a government anti-poverty effort.

It is worth noting that NREGA was designed with the intent to both lower poverty and increase female empowerment by increasing women's labor force participation and earnings potential. These results suggest, however, that it could be unintentionally decreasing the future earnings potential of its beneficiaries by inducing them to drop out of school earlier than they otherwise would have. This is especially true for girls, who, rather than at least gaining market experience and their own earnings like their male counterparts, are simply substituting for their mothers at home.

This research fits into the larger literature which attempts to document the (sometimes unanticipated) price effects of poverty alleviation programs, both in the United States (such as Hastings and Washington (2010) and Rothstein (2010)) and in developing countries. This includes both the extensive literature on conditional and unconditional cash transfers (see Attanasio *et al.* (2011), among many others), as well as those programs with undesigned, negative price implications (such as Jayachandran *et al.* (2013) and Kablonski and Townsend (2011)).

While the policy implications of this paper might seem straightforward, it's important to remember that this analysis represents the effect of NREGA on one particular outcome that may be of interest to policy makers (human capital investment). While we would argue that this is quite an important outcome, we are not measuring the benefits that the program provides in terms of consumption, protection against income shocks, or any number of other outcomes. Thus, we are not in a position to measure the overall welfare impact of this particular anti-poverty program.

Rather, the takeaway from these results is that social programs have price effects, and that these price effects can have very real consequences. It is important for these price effects to be understood so that social programs can be designed in order to maximize their potential to increase economic growth and alleviate poverty.

References

Akresh, R., Bagby, E., de Walque, D. and Kazianga, H. (2010). Child ability and household human capital investment decisions in burkina faso, iZA Discussion Paper 5326.

Alderman, H., Orazem, P. and Paterno, E. (2001). School quality, school cost, and the public/private school choices of low-income households in pakistan. *Journal of Human Resources*, **36** (2), 304–326.

Almond, D. (2006). Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 u.s. population. *Journal of Political Economy*, **114** (4), 672ñ712.

ASER (2007). *Annual Status of Education Report (Rural) 2007*. Tech. rep., ASER.

— (2010). *Annual Status of Education Report (Rural) 2009*. Annual report, Pratham.

— (2013). *Sample Design of Rural ASER*. Tech. rep., ASER.

Atkin, D. (2009). Endogenous skill acquisition and export manufacturing in mexico. *mimeo, Yale University*.

— (2012). Endogenous skill acquisition and export manufacturing in mexico, working Paper.

Attanasio, O. P., Meghir, C. and Santiago, A. (2011). Education choices in mexico: Using a structural model and a randomized experiment to evaluate progresa. *Review of Economic Studies*, **79**, 37–66.

Azam, M. (2012). The impact of indian job guarantee scheme on labor market outcomes: Evidence from a natural experiment, iZA Working Paper 6548.

Bandyopadhyay, S., Kanji, S. and Wang, L. (2012). The impact of rainfall and temperature variation on diarrheal prevalence in sub-saharan africa. *Applied Geography*, **33** (0), 63 – 72, the Health Impacts of Global Climate Change: A Geographic Perspective.

Banerjee, A. and Somanathan, R. (2007). The political economy of public goods: Some evidence from india. *Journal of Development Economics*, **82**, 287ñ314.

Basu, A. K., Chau, N. H. and Kanbur, R. (2009). A theory of employment guarantees: Contestability credibility and distributional concerns. *Journal of Public Economics*, **93** (3-4), 482497.

Berg, E., Bhattacharyya, S., Durgam, R. and Ramachandra, M. (2012). Can rural public works affect agricultural wages? evidence from lndia, cSAE Working Paper.

Burde, D. and Linden, L. L. (2009). The effect of proximate schools: A randomized controlled trial in afghanistan. *Mimeo, Columbia University*.

Carlton, E. J., Eisenberg, J. N. S., Goldstick, J., Cevallos, W., Trostle, J. and Levy, K. (2013). Heavy rainfall events and diarrhea incidence: The role of social and environmental factors. *American Journal of Epidemiology*.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. and Rogers, F. H. (2006). Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives*, **20** (1), 91–116.

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

Curriero, F., Patz, J., Rose, J. and Lele, S. (2001). The association between extreme precipitation and waterborne disease outbreaks in the united states, 1948ñ1994. *American Journal of Public Health*, **91** (8), 1194ñ1199.

Dossani, R. and Kenney, M. (2004). The next wave of globalization? exploring the relocation of service provision to india. *Working Paper 156*.

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review*, **91** (4).

—, Hanna, R. and Ryan, S. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, **102** (4), 1241–1278.

Duryea, S. and Arends-Kuenning, M. (2003). School attendance, child labor and local labor market fluctuations in urban brazil. *World Development*, **31** (7), 1165–1178.

Evans, D., Kremer, M. and Ngatia, M. (2008). The impact of distributing school uniforms on children's education in kenya. *mimeo, World Bank*.

Foster, A. and Rosenzweig, M. (1996). Technical change and human capital returns and investments: Evidence from the green revolution. *American Economic Review*, **86** (4), 931–953.

Freeman, R. (1976). *The Overeducated American*. New York: Academic Press.

Funkhouser, E. (1999). Cyclical economic conditions and school attendance in costa rica. *Economics of Education Review*, **18** (1), 31–50.

Giridharadas, A. (December 30, 2010). The caste buster. *New York Times*.

Glewwe, P. and Jacoby, H. G. (2004). Economic growth and the demand for education: is there a wealth effect? *Journal of Development Economics*, **74** (1), 33–51.

Glick, P. and Sahn, D. E. (2000). Schooling of girls and boys in a west african country: the effects of parental education, income, and household structure. *Economics of Education Review*, **19**, 63–87.

Government of India (2011). Census 2011.

Griliches, Z. (1997). Education, human capital, and growth: A personal perspective. *Journal of Labor Economics*, **15**, S330–S344.

Haque, U., Hashizume, M., Glass, G. E., Dewan, A. M., Overgaard, H. J. and Yamamoto, T. (2010). The role of climate variability in the spread of malaria in bangladeshi highlands. *PLoS ONE*, **5** (12).

Hastings, J. and Washington, E. (2010). The first of the month effect: Consumer behavior and store responses. *American Economic Journal: Economic Policy*, **2**, 142–162.

Heath, R. and Mobarak, M. (2012). Does demand or supply constrain investments in education? evidence from garment sector jobs in bangladesh. *Mimeo, Yale University*.

Heckman, J. (1993). *Private Sector Skill Formation: International Comparisons*, University of Chicago Press, chap. Determinants of Young Male Schooling and Training Choices.

Imbert, C. and Papp, J. (2012a). Equilibrium distributional impacts of government employment programs: Evidence from indiaís employment guarantee, paris School of Economics Working Paper.

— and — (2012b). Equilibrium distributional impacts of government employment programs: Evidence from indias employment guarantee, paris School of Economics Working Paper.

— and — (2014). Short-term migration, rural workfare programs and urban labor markets: Evidence from india, mimeo.

IPCC (2007). *Summary for policymakers. Climate Change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Jacoby, H. G. and Skoufias, E. (1997). Risk, financial markets, and human capital in a developing country. *The Review of Economic Studies*, **64** (3), 311–335.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy*, **114** (3).

—, Cunha, J. M. and de Giorgi, G. (2013). The price effects of cash versus in-kind transfers.

Jensen, R. (2000). Agricultural volatility and investments in children. *The American Economic Review*, **90** (2), 399–404.

— (2010). The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics*, **125**, 515–548.

— (2012). Do labor market opportunities affect young women's work and family decisions? experimental evidence from india. *Quarterly Journal of Economics*, **127** (2), 753–792.

— and Oster, E. (2009). The power of tv: Cable television and women's status in india. *Quarterly Journal of Economics*, **124** (3), 465–476.

Kablonski, J. P. and Townsend, R. M. (2011). A structural evaluation of a large-scale quasi-experimental microfinance initiative. *Econometrica*, **79** (5), 1357–1406.

Kane, T. (1994). College entry by blacks since 1970: The role of college costs, family background and the returns to education. *Journal of Political Economy*, **102** (5), 878–911.

Katz, L. and Murphy, K. (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics*, **107** (1), 35–78.

Kaur, S. (2011). Nominal wage rigidity in village labor markets, harvard University Working Paper.

Kochar, A. (1999). Smoothing consumption by smoothing income: Hours-of-work responses to idiosyncratic agricultural shocks in rural india. *The Review of Economics and Statistics*, **81** (1), 50–61.

Kremer, M. (2003). Randomized evaluations of educational programs in developing countries: Some lessons. *American Economic Review*, **92** (2).

—, Chaudhury, N., Rogers, H., Muralidharan, K. and Hammer, J. (2005). Teacher absence in india: A snapshot. *Journal of the European Economic Association*, **3** (2-3).

Kruger, D. (2007). Coffee production effects on child labor and schooling in rural brazil. *Journal of Development Economics*, **82**, 448ñ463.

Kumar, A., Valecha, N., Jain, T. and Dash, A. P. (2007). Burden of malaria in india: Retrospective and prospective view. *American Journal of Tropical Medicine and Hygiene*, **77**, 69ñ78.

LaFerrara, E., Chong, A. and Duryea, S. (2009). Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics*.

Li, T. and Sehkri, S. (2013). The unintended consequences of employment based safety net programs, mimeo.

Maccini, S. and Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review*, **99** (3), 1006–26.

Mahajan, V. and Gupta, R. K. (2011). Non farm opportunities for smallholder agriculture, paper presented at the IFAD Conference on New Directions for Smallholder Agriculture.

Mankiw, N. G., Romer, D. and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, **107** (2), pp. 407–437.

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

Miller, G. and Urdinola, B. P. (2010). Cyclicality, mortality, and the value of time: The case of coffee price fluctuations and child survival in colombia. *Journal of Political Economy*, **118** (1).

Ministry of Rural Development (2005). *Mahatma Gandhi National Rural Employment Guarantee Act (Mahatma Gandhi NREGA)*. Tech. rep., Ministry of Rural Development, Government of India.

Ministry of Rural Development (2008). *The National Rural Employment Guarantee Act 2005 (NREGA)*. Tech. rep., Ministry of Rural Development, Government of India.

Munshi, K. and Rosenzweig, M. (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, **96** (4).

— and — (2009). Why is mobility in india so low? social insurance, inequality and growth, nBER Working Paper No. 14850.

NASSCOM (2010). Strategic review 2010. *New Delhi: National Association of Software and Service Companies*.

Ng, C. and Mitter, S. (2005). Valuing women's voices: Call center workers in malaysia and india. *Gender and Technology Development*, **9**, 209–233.

Orazem, P. and King, E. M. (2007). Schooling in developing countries: The roles of supply, demand and government policy. *Iowa State University, Department of Economics Working Paper 07019*.

Pathania, V. (2007). The long run impact of drought at birth on height of women in rural india, working Paper.

Rao, C. H. H., Ray, S. K. and Subbarao, K. (1988). *Unstable Agriculture and Droughts : Implications for Policy*. New Delhi: Vikas Publishing House.

Ravallion, M. (1987). Market responses to anti-hunger policies: Effects on wages, prices and employment, wIDER Working Paper 28.

Rose, E. (2001). Ex ante and ex post labor supply response to risk in a low-income area. *Journal of Development Economics*, **64** (2), 371–388.

Rosenzweig, M. and Evenson, R. (1977). Fertility, schooling, and the economic contribution of children in rural india: An econometric analysis. *Econometrica*, **45** (5), 1065–1079.

Rothstein, J. (2010). Is the eitc as good as an nit? conditional cash transfers and tax incidence. *American Economic Journal: Economic Policy*, **2**, 177–208.

Schady, N. (2004). Do macroeconomic crises always slow human capital accumulation? *World Bank Economic Review*, **18** (2), 131–154.

Schultz, P. (2004). School subsidies for the poor: Evaluating the mexican progresa poverty program. *Journal of Development Economics*, **74** (1), 199–250.

Shah, M. and Steinberg, B. (2014). Do workfare programs decrease human capital investments? evidence from nrega in india, working Paper.

Shastry, G. K. (2012). Human capital response to globalization: Education and information technology in india. *Journal of Human Resources*, **47** (2), 287–330.

Singh, A., Park, A. and Dercon, S. (forthcoming). School meals as a safety net: An evaluation of the midday meal scheme in india. *Economic Development and Cultural Change*.

Thomas, D., Beegle, K., Frankenberg, E., Sikoki, B., Strauss, J. and Teruel, G. (2004). Education in a crisis. *Journal of Development Economics*, **74** (1), 53–85.

Topalova, P. (2005). Trade liberalization, poverty, and inequality: Evidence from indian districts, nBER Working Paper No. 11614.

Vagh, S. B. (2009). Validating the aser testing tools: Comparisons with reading fluency measures and the read india measures, aSER Centre Working Paper.

Zimmerman, L. (2014). Why guarantee employment? evidence from a large indian publicworks program, mimeo.

Appendix A

Appendix to Chapter 1

A.1 Appendix Tables

A.1.1 Calculating ITES Center Impact on Income

In the text we report an estimated percentage increase in income due to ITES centers. We use a simple back-of-the-envelope calculation to generate this number. Although this is very unlikely to be perfectly accurate, we argue it is an upper bound on what the impact might be. To generate this number, we focus on the median PIN code and the median ITES center (in terms of size). We observe the number of children enrolled in school in the median PIN code; this number is roughly 10,000. We use this number to estimate the average number of people in the median PIN code. To do this, we note that the Indian census indicates children aged 6-12 make up roughly 15% of the overall population, and from the National Family and Health Survey (which is consistent with other sources) we observe that roughly 85% of children in this age range are enrolled in school. Combining these figures we argue that roughly 12.75% of individuals are children enrolled in school, so we expect the median PIN code to have roughly 78,000 people.

The median per capita income in these states is US$659. Applying this value to the population, we estimate total income of the median PIN code at about US$51 million.

The median ITES center in our sample has 80 employees and based on a survey of a sub-sample of ITES centers, pays roughly US$2100 per year in starting wages. This is more than twice the median per capita income. We calculate the increase in income due to ITES centers assuming that the income from ITES center employees is simply added to total income in the PIN code; we note this is likely to be an *overestimate*, since these individuals probably substitute into ITES center jobs away from some lower paying job, not from doing nothing.

We calculate the increased income, and then calculate the percentage increase implied by this; the resulting figure is 0.57% as reported.

A.1.2 Madurai Survey

We conducted a survey of approximately 1,000 households in Madurai district in Tamil Nadu. Of these, 500 households were in the city of Madurai, 250 were in Thirumangalam, a town approximately 20 kilometers away, and 250 were in Peraiyur, a town approximately 50 kilometers away. We surveyed households in groups of 10: at the start, 100 households were randomly selected from election rolls. These 100 households were each surveyed along with their 9 closest neighbors.

The survey included a household roster, with the names of each member of the household (including those who did not live at home), along with age, highest grade completed, enrollment status, employment status, job, and distance to work. In addition, we asked the household questions about assets, language, and how long they have lived in the area. We also asked questions about earnings for individuals with primary school and secondary school in the area, as well as a series of questions about ITES centers. This latter set of questions included information on whether the individual knew anyone who worked at one of these centers, whether they knew of any of these businesses in the local area and a series of questions about what qualifications were required for this job. If households reported knowing of a BPO in the local area, we included an open response question about whether they had made any changes because of the ITES center.

If the household included at least one child between the ages of 5 and 15 and enrolled in school, we randomly selected one of the children for more detailed questions about schooling and future job and marriage. These included questions on type of school attended (e.g. public, private) and language of instruction. We asked parents to choose the three most likely jobs for their child to have from a list of 15. We asked these questions before any mention of BPOs or ITES centers, in order to avoid leading the respondents in any way. All schooling (and other) questions were asked of the head of household, typically the father.

In addition to this, we recorded the GPS location of each household surveyed, as well as the GPS locations of BPOs in the local area in order to accurately calculate distance from the nearest ITES center. Since all of the BPOs in the district were located in Madurai city,

Table A.1: *Madurai Survey Summary Statistics*

	Mean	Std. Dev.	Observations
Information:			
Heard of someone who has worked at an ITES (share)	.207	.405	995
Knows of an ITES center in the local area (share)	.137	.344	996
% ITES Ques. "Don't Know"	.172	.246	1000
% True Qual. Answer Correct	.894	.239	1000
% False Qual. Answer Correct	.411	.316	1000
Expectations:			
BPO listed as first job for child	.045	.208	398
Estimated return to secondary school	818.15	1911.2	1000

the households in Peraiyur and Thirumangalam were necessarily at least 20 kilometers away from any ITES center, but within the city of Madurai there was significant variance in distance to the nearest BPO.

A.1.3 Appendix Figures

Figure A.1: *Distance to Neariest ITES Center and Knowing Someone who Works in One*

A.1.4 Appendix Tables

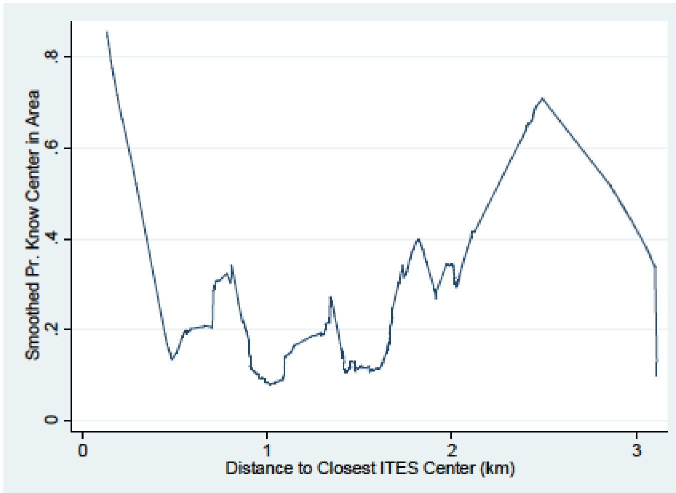


Figure A.2: *Distance to Neariest ITES Center and Knowing Center in Local Area*

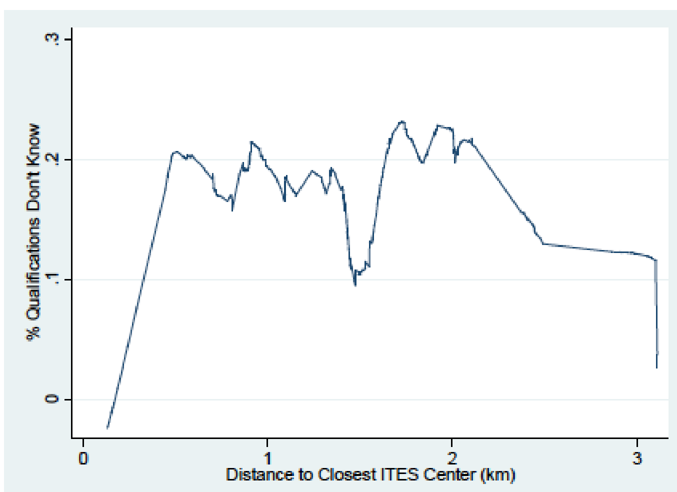


Figure A.3: *Distance to Neariest ITES Center and % of Qualifications Report "Don't Know"*

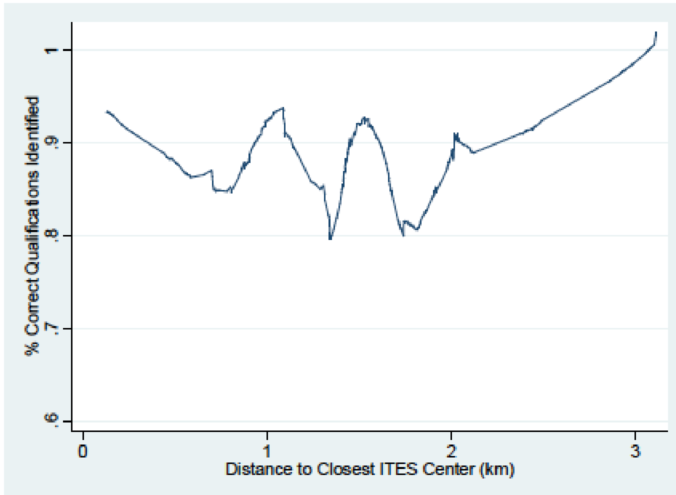


Figure A.4: *Distance to Neariest ITES Center and % of Correct Qualifications Identified*

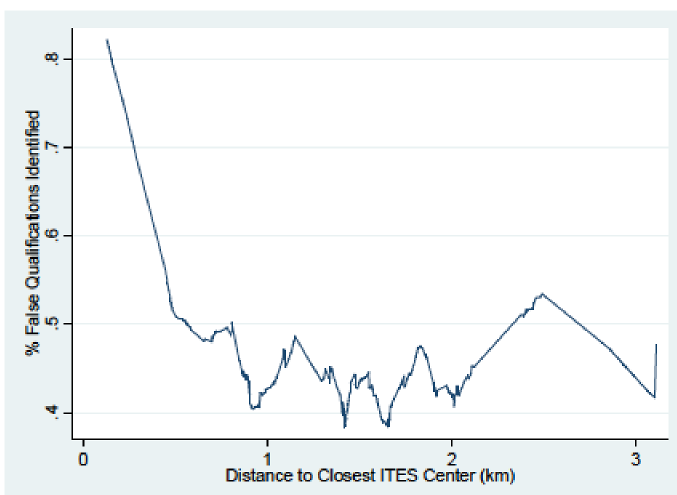


Figure A.5: *Distance to Neariest ITES Center and % of False Qualifications Identified as False*

Table A.2: *Enrollment Effects by Demographic Group, State*

Dependent Variable:		*Log Enrollment*		
		Number of ITES Centers	Standard Error	Observations
(1)	Boy Enrollment	.048**	.018	896,078
(2)	Girl Enrollment	.046***	.014	900,899
(3)	Grades 1-2 Enroll.	.032	.024	822,569
(4)	Grades 3-4 Enroll.	.034	.021	810,926
(5)	Grades 5-6 Enroll.	.046**	.020	851,483
(6)	Grades 7-8 Enroll.	.063**	.028	338,995
(7)	Andra Pradesh	.023	.042	358,934
(8)	Karnataka	.078**	.038	311,704
(9)	Tamil Nadu	.048***	.013	240,861

Notes: This table shows coefficients on number of ITES centers from regression of the form in Column 1 of Table 4 but with variation in the left hand side variable. All regressions include the standard controls: School fixed effects, time-varying school plant characteristics, year dummies interacted with dummies for state, urban, school language of instruction and English language school in village. Standard errors in parentheses, clustered at the neighborhood level. *significant at 10% **significant at 5% ***significant at 1%.

Table A.3: *Robustness Checks for Main Results*

Panel A: First Differences in Number of ITES Centers				
Sample:	All Schools		In PIN Code with any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
Change in # ITES Centers	.037*	.037**	.042*	.074**
	(.021)	(.018)	(.023)	(.029)
Panel B: District Trends in All Specifications				
Sample:			In PIN Code with any English-Language Schools	Ever Had an ITES Center
Controls:			*District Trends*	*District Trends*
# ITES Centers			.027*	.054***
			(.014)	(.017)
Panel C: Levels of Enrollment				
Sample:	All Schools		In PIN Code with any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
# ITES Centers	22.0**	19.0***	21.9*	38.4***
	(10.7)	(3.5)	(11.4)	(12.7)
Panel D: Effect of Any ITES Center				
Sample:	All Schools		In PIN Code with any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
Any ITES Center	.029	.005	.032	.030
	(.024)	(.014)	(.027)	(.034)
Panel E: Limit to PIN Codes with ITES Centers Added in 2005/2006				
Sample:			PIN Codes Which Added ITES Centers in 2005 or 2006 only	
Controls:			*Standard*	
# ITES Centers			.090**	
			(.041)	

Notes: This table shows several robustness checks for our main specification in Table 3. Panel A shows the effect of first differences in the number of ITES centers on log enrollment. Panel B shows the effect of the number of ITES centers on log enrollment, with district trends included for all specifications. Panel C shows the effect of the number of ITES centers on total enrollment. Panel D shows the effect of any ITES center in the PIN code on log enrollment. Panel E shows the effect of the number of ITES centers on log enrollment for those PIN codes which added call centers in 2005 and 2006. Standard errors in parentheses, clustered at the PIN code level. *significant at 10% **significant at 5% ***significant at 1%.

Table A.4: *Trends Leading Up to ITES Center Entry*

Dependent Variable:	*Log Enrollment*			
Sample:	All Schools		In PIN Code with Any English Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
	(1)	(2)	(3)	(4)
ITES Centers	.070***	.056**	.070***	.080***
	(.026)	(.019)	(.025)	(.029)
Years To Entry	.008	.008	.008	.003
	(.007)	(.005)	(.008)	(.010)
Observations	911,499	911,499	314,476	2,121

Standard controls: School fixed effects, time-varying school plant characteristics, year dummies interacted with dummies for state, urban, school language, and English language school in PIN code.
District Trend Controls: Standard controls plus district-specific trends.

Notes: This table provides a second test for pretrends in enrollment. The independent variable measures the number of ITES centers in the same PIN code as the school, along with the linear trend in enrollment leading up to the entry of an ITES center. Columns 1-2 include all schools. Column 3 is limited to PIN codes with any English schools. Column 4 is limited to schools which ever have an ITES center in their PIN code (either always have the same number or change during the sample). Standard errors (in parentheses) are clustered at the PIN code level in Columns 1, 3 and 4; clustered errors could not be estimated when district trends are included in Column 2. *significant at 10% **significant at 5% ***significant at 1%. All regressions are weighted by initial school enrollment level.

Table A.5: *Effects on Enrollment with Population Controls*

Dependent Variable:	*Log Enrollment*			
Panel A: Restricted Sample, No Population Control				
Sample:	All Schools		In PIN Code with Any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
	(1)	(2)	(3)	(4)
Number of ITES Centers	.061**	.053***	.073**	.094**
	(.029)	(.013)	(.033)	(.042)
Observations	327,144	327,144	127,911	1,073
Panel B: Restricted Sample, with Population Control				
Sample:	All Schools		In PIN Code with Any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
	(1)	(2)	(3)	(4)
Number of ITES Centers	.060**	.053***	.071**	.093**
	(.029)	(.013)	(.032)	(.038)
Log Village Population	.003	.003***	.003	-.013
	(.002)	(.0007)	(.004)	(.011)
Observations	327,144	327,144	127,911	1,073

Standard controls: School fixed effects, time-varying school plant characteristics, year dummies interacted with dummies for state, urban, school language, and English language school in village. District Trend controls: Standard controls plus district-specific trends.

Notes: This table shows the impact of ITES centers controlling for population. Population is reported by a subset of school-years, and is reported by the school as the village population. In cases where the school does not report population but other schools in the village do report population, we use the average population among reporter schools as population for all schools in the village. Panel A does not control for population but limits the sample to school-years in which population is observed. Standard errors (in parentheses) are clustered at the village level. *significant at 10% **significant at 5% ***significant at 1%.

Table A.6: *Effect of ITES Centers on Schools*

Panel A: Number of Schools				
Dependent Variable:	*Count of Schools in Village*			
Sample:	All Schools		In PIN Code with Any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
Number of ITES Centers	-.013	-.007	-.019	-.049
	(.048)	(.019)	(.100)	(.080)
Observations	356,796	356,796	41,901	477
Panel B: Medium of Instruction				
Dependent Variable:	*Number of Schools with English Instruction*			
Sample:	All Schools		In PIN Code with Any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
Number of ITES Centers	.001	.001	.001	-.001
	(.003)	(.003)	(.004)	(.002)
Observations	977,543	977,543	338,022	2,416
Panel C: School Infrastructure				
Dependent Variable:	*Composite Infrastructure Measure*			
Sample:	All Schools		In PIN Code with Any English-Language Schools	Ever Had an ITES Center
Controls:	*Standard*	*District Trends*	*Standard*	*Standard*
Number of ITES Centers	-.022	-.044	-.049	-.145***
	(.052)	(.028)	(.049)	(.045)
Observations	990,540	990,540	344,479	2,416

Standard controls: Village fixed effects, time-varying school plant characteristics, year dummies interacted with dummies for state, urban, and English language school in village.
District Trend controls: Standard controls plus district-specific trends.

Notes: This table shows the effect of ITES centers on schools in the PIN Code. Panel A shows the effect of ITES centers on the number of schools in a village. Panel B shows the effect of ITES centers on the number of schools with English instruction. Panel C shows the effect of ITES centers on a composite measure of school infrastructure (the first principle component of an index of electricity, toilets, classroom quality and boundary wall). Columns 1-2 include all schools. Column 3 is limited to villages with any English schools. Column 4 is limited to schools which ever have an ITES center in their PIN code (either always have the same number or change during the sample). Standard errors (in parentheses) are clustered at the village level in Columns 1, 3 and 4; clustered errors could not be estimated when district trends are included in Column 2. *significant at 10% **significant at 5% ***significant at 1%.

Appendix B

Appendix to Chapter 2

B.1 Mathematical Appendix

Parents maximize total household utility:

$$\max_{c_2, s_2} \left\{ u_1(c_1) + \beta u_2(c_2) + \beta^2 V(e_3) \right\}$$

subject to the constraints that

$$c_2 \leq w_2(h + (1 - s_2)e_2)$$

$$c_1 \leq w_1 h$$

$$s_2 \in [0, 1]$$

$$e_3 = f_3\left(e_2, c_2, s_2\right) = f_3\left(e_2, w_2(h + (1 - s_2)e_2), s_2\right)$$

and we have assumed that

$$V(e_3) = e_3$$

Since no decisions are made in the first period, we can rewrite the maximization problem as

$$\max_{s_2 \in [0,1], c_2} \{u_2(c_2) + \beta f_3(e_2, c_2, s_2)\} \text{ s.t. } c_2 \le w_2(h + (1 - s_2)e_2)$$

The Lagrangian for this problem is:

$$\mathcal{L} = u_2(c_2) + \beta f_3(e_2, s_2, c_2) + \lambda [w_2(h + (1 - s_2)e_2) - c_2]$$

Given the assumptions laid out in Section 2.2, the first order conditions will be necessary and sufficient to characterize a unique optimum:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \lambda} &= w_2(h + (1 - s_2^*(w_2, e_2)e_2)) - c_2^*(w_2, e_2) = 0 \\
\frac{\partial \mathcal{L}}{\partial c_2} &= u'(c_2^*(w_2, e_2)) + \beta \frac{\partial f_3}{\partial c_2}(e_2, c_2^*(w_2, e_2), s_2^*(w_2, e_2)) - \lambda^*(w_2, e_2) = 0 \\
\frac{\partial \mathcal{L}}{\partial s_2} &= \beta \frac{\partial f_3}{\partial s_2}(e_2, c_2^*(w_2, e_2), s_2^*(w_2, e_2)) - \lambda^*(w_2, e_2) \cdot w_2 e_2 = 0
\end{aligned}$$

Rearranging yields the single first-order condition outlined in the text of Section 2:

$$w_2 e_2 \frac{\partial u_2}{\partial c_2} = \beta \left(\frac{\partial f_3}{\partial s_2} - w_2 e_2 \frac{\partial f_3}{\partial c_2} \right)$$

We define the function

$$\Theta(w_2, e_2, c_2^*, s_2^*) = \left(\frac{\partial f_3}{\partial s_2} - w_2 e_2 \frac{\partial f_3}{\partial c_2} \right)$$

so we can rewrite the first order condition as

$$w_2 e_2 \frac{\partial u_2}{\partial c_2} = \beta \Theta$$

Effect of School-Aged Wages on Schooling and Human Capital

Beginning from the first order condition,

$$\text{F.O.C.:}\;\; w_2 e_2 \frac{\partial u_2}{\partial c_2} = \beta\left(\frac{\partial f_3}{\partial s_2} - w_2 e_2 \frac{\partial f_3}{\partial c_2}\right)$$

We can take the total derivative with respect to w_2 (noting that $e_2 = f_2(w_1 h)$ and $e_3 = f_3(w_1 h, w_2(h + (1 - s_2)e_2), s_2)$:

$$\left[\frac{d^2 u_2}{dc_2^2} w_2 e_2 \left(\left(h + \left(1 - s_2^*\right) e_2\right) - w_2 e_2 \frac{\partial s_2^*}{\partial w_2}\right) + e_2 \frac{du_2}{dc_2}\right] =$$

$$\beta\left[\frac{\partial^2 f_3}{\partial s_2 \partial c_2}\left(\left(h + \left(1 - s_2^*\right) e_2\right) - w_2 e_2 \frac{\partial s_2^*}{\partial w_2}\right) + \frac{\partial^2 f_3}{\partial s_2^2}\frac{\partial s_2^*}{\partial w_2}\right]$$

$$-\beta\left[\left(\frac{\partial^2 f_3}{\partial c_2^2}\left(\left(h + \left(1 - s_2^*\right) e_2\right) - w_2 e_2 \frac{\partial s_2^*}{\partial w_2}\right) + \frac{\partial^2 f_3}{\partial c_2 \partial s_2}\frac{\partial s_2^*}{\partial w_2}\right) e_2 w_2 + e_2 \frac{\partial f_3}{\partial c_2}\right]$$

and solve for $\frac{\partial s_2^*}{\partial w_2}$

$$\frac{\partial s_2^*}{\partial w_2} = \frac{\frac{d^2 u_2}{dc_2^2} w_2 e_2 \left(h + \left(1 - s_2^*\right) e_2\right) + e_2 \frac{du_2}{dc_2} - \beta\left[\frac{\partial^2 f_3}{\partial s_2 \partial c_2}\left(h + \left(1 - s_2^*\right) e_2\right) - \frac{\partial^2 f_3}{\partial c_2^2}\left(h + \left(1 - s_2^*\right) e_2\right) e_2 w_2 - e_2 \frac{\partial f_3}{\partial c_2}\right]}{w_2^2 e_2^2 \frac{d^2 u_2}{dc_2^2} + \beta\left[\frac{\partial^2 f_3}{\partial s_2^2} + w_2^2 e_2^2 \frac{\partial^2 f_3}{\partial c_2^2} - 2 e_2 w_2 \frac{\partial^2 f_3}{\partial c_2 \partial s_2}\right]}$$

Since, by assumption

$$w_2^2 e_2^2 \frac{d^2 u_2}{dc_2^2} + \beta\left[\frac{\partial^2 f_3}{\partial s_2^2} + w_2^2 e_2^2 \frac{\partial^2 f_3}{\partial c_2^2} - 2 e_2 w_2 \frac{\partial^2 f_3}{\partial c_2 \partial s_2}\right] < 0$$

we can write that

$$\frac{\partial s_2^*}{\partial w_2} \propto -\frac{d^2 u_2}{dc_2^2} w_2 e_2 \left(h + \left(1 - s_2^*\right) e_2\right) - e_2 \frac{du_2}{dc_2}$$

$$+\beta\left[\frac{\partial^2 f_3}{\partial s_2 \partial c_2}\left(h + \left(1 - s_2^*\right) e_2\right) - \frac{\partial^2 f_3}{\partial c_2^2}\left(h + \left(1 - s_2^*\right) e_2\right) e_2 w_2 - e_2 \frac{\partial f_3}{\partial c_2}\right]$$

Rearranging slightly,

$$\frac{\partial s_2^*}{\partial w_2} \propto -e_2\left(\frac{du_2}{dc_2} + \beta \frac{\partial f_3}{\partial c_2}\right) - \left(h + \left(1 - s_2^*\right) e_2\right) w_2 e_2 \frac{d^2 u_2}{dc_2^2} + \left(h + \left(1 - s_2^*\right) e_2\right) \beta\left[\frac{\partial^2 f_3}{\partial s_2 \partial c_2} - \frac{\partial^2 f_3}{\partial c_2^2} e_2 w_2\right]$$

And since

$$\Theta = \frac{\partial f_3}{\partial s_2} - w_2 e_2 \frac{\partial f_3}{\partial c_2}$$

$$\frac{\partial \Theta}{\partial c_2} = \frac{\partial^2 f_3}{\partial s_2 \partial c_2} - e_2 w_2 \frac{\partial^2 f_3}{\partial c_2^2}$$

And substituting this into the above equation yields

$$\frac{\partial s_2^*}{\partial w_2} \propto -e_2 \left(\frac{du_2}{dc_2} + \beta \frac{\partial f_3}{\partial c_2} \right) - \left(h + \left(1 - s_2^*\right) e_2\right) w_2 e_2 \frac{d^2 u_2}{dc_2^2} + \left(h + \left(1 - s_2^*\right) e_2\right) \beta \frac{\partial \Theta}{\partial c_2}$$

In addition, since we know that

$$e_3^* = f_3\left(e_2, c_2^*, s_2^*\right)$$

we can write the effect of school-aged wages on later-life human capital:

$$\begin{aligned}
\frac{d}{dw_2}\left(f_3\left(e_2, c_2^*, s_2^*\right)\right) &= \frac{\partial f_3}{\partial c_2}\frac{\partial c_2^*}{\partial w_2} + \frac{\partial f_3}{\partial s_2}\frac{\partial s_2^*}{\partial w_2} \\
&= \frac{\partial f_3}{\partial c_2}\left(h + (1 - s_2^*)e_2 - w_2 e_2 \frac{\partial s_2^*}{\partial w_2}\right) + \frac{\partial f_3}{\partial s_2}\frac{\partial s_2^*}{\partial w_2} \\
&= \left(h + (1 - s_2)e_2\right)\frac{\partial f_3}{\partial c_2} + \Theta \frac{\partial s_2^*}{\partial w_2}
\end{aligned}$$

Effect of Early Life Wages on Schooling and Human Capital

Beginning from the first order condition,

$$\text{F.O.C.:}\ \frac{\partial u_2}{\partial c_2} w_2 e_2 = \beta \left(\frac{\partial f_3}{\partial s_2} - w_2 e_2 \frac{\partial f_3}{\partial c_2} \right)$$

Noting as in the text, that

$$\frac{d}{dw_1}(e_2) = \frac{\partial f_2}{\partial c_1}\frac{\partial c_1}{\partial w_1} = h\frac{\partial f_2}{\partial c_1} > 0$$

an thus, in order to understand the effect of early life wages on schooling and later-life human capital, it is sufficient to study the effect of period 2 human capital on these variables:

$$\frac{\partial s_2^*}{\partial w_1} = h\frac{\partial f_2}{\partial c_1}\frac{\partial s_2^*}{\partial e_2} \propto \frac{\partial s_2^*}{\partial e_2}$$

We can take the total derivative of the first order condition with respect to e_2 (noting that $e_2 = f_2\left(w_1 h\right)$ and $e_3 = f_3\left(w_1 h, w_2(h+(1-s_2)e_2), s_2\right)$:

$$\begin{gathered}
w_2 e_2 \left(w_2\left(1-s_2^*\right) - w_2 e_2 \frac{\partial s_2^*}{\partial e_2}\right)\frac{\partial^2 u_2}{\partial c_2^2} + w_2\frac{\partial u_2}{\partial c_2} = \\
\beta\left[\frac{\partial^2 f_3}{\partial s_2 \partial e_2} + \frac{\partial^2 f_3}{\partial s_2 \partial c_2}\left(w_2\left(1-s_2^*\right) - w_2 e_2\frac{\partial s_2^*}{\partial e_2}\right) + \frac{\partial^2 f_3}{\partial s_2^2}\frac{\partial s_2^*}{\partial e_2}\right] \\
-\beta\left[w_2 e_2\left(\frac{\partial^2 f_3}{\partial c_2 \partial e_2} + \frac{\partial^2 f_3}{\partial c_2^2}\left(w_2\left(1-s_2^*\right) - w_2 e_2\frac{\partial s_2^*}{\partial e_2}\right) + \frac{\partial^2 f_3}{\partial c_2 \partial s_2}\frac{\partial s_2^*}{\partial e_2}\right) + w_2\frac{\partial f_3}{\partial c_2}\right]
\end{gathered}$$

Rearranging, we have

$$\frac{\partial s_2^*}{\partial e_2} = \frac{-w_2^2 e_2\left(1-s_2^*\right)\frac{\partial^2 u_2}{\partial c_2^2} - w_2\frac{\partial u_2}{\partial c_2} + \beta\left[\frac{\partial^2 f_3}{\partial s_2 \partial e_2} + w_2\left(1-s_2^*\right)\left(\frac{\partial^2 f_3}{\partial s_2 \partial c_2} - w_2 e_2\frac{\partial^2 f_3}{\partial c_2^2}\right) - w_2 e_2\frac{\partial^2 f_3}{\partial c_2 \partial e_2} - w_2\frac{\partial f_3}{\partial c_2}\right]}{-w_2^2 e_2^2\frac{\partial^2 u_2}{\partial c_2^2} - \beta\left[\frac{\partial^2 f_3}{\partial s_2^2} - 2w_2 e_2\frac{\partial^2 f_3}{\partial s_2 \partial c_2} + w_2^2 e_2^2\frac{\partial^2 f_3}{\partial c_2^2}\right]}$$

and since, by assumption,

$$-w_2^2 e_2^2\frac{\partial^2 u_2}{\partial c_2^2} - \beta\left[\frac{\partial^2 f_3}{\partial s_2^2} - 2w_2 e_2\frac{\partial^2 f_3}{\partial s_2 \partial c_2} + w_2^2 e_2^2\frac{\partial^2 f_3}{\partial c_2^2}\right] > 0$$

$$\begin{gathered}
\frac{\partial s_2^*}{\partial e_2} \propto -w_2^2 e_2\left(1-s_2^*\right)\frac{\partial^2 u_2}{\partial c_2^2} - w_2\frac{\partial u_2}{\partial c_2} + \\
\beta\left[\frac{\partial^2 f_3}{\partial s_2 \partial e_2} + w_2\left(1-s_2^*\right)\left(\frac{\partial^2 f_3}{\partial s_2 \partial c_2} - w_2 e_2\frac{\partial^2 f_3}{\partial c_2^2}\right) - w_2 e_2\frac{\partial^2 f_3}{\partial c_2 \partial e_2} - w_2\frac{\partial f_3}{\partial c_2}\right]
\end{gathered}$$

and since

$$\Theta = \frac{\partial f_3}{\partial s_2} - w_2 e_2\frac{\partial f_3}{\partial c_2}$$

we can rewrite this expression as

$$\frac{\partial s_2^*}{\partial e_2} \propto -w_2^2 e_2 \left(1 - s_2^*\right) \frac{\partial^2 u_2}{\partial c_2^2} - w_2 \frac{\partial u_2}{\partial c_2} + \beta \left[\frac{\partial \Theta}{\partial e_2} + w_2 \left(1 - s_2^*\right) \frac{\partial \Theta}{\partial c_2}\right]$$

In addition, since we know that

$$e_3^* = f_3 \left(e_2, c_2^*, s_2^*\right)$$

we can write the effect of school-aged human capital on later-life human capital:

$$\begin{aligned}
\frac{d}{de_2}\left(e_3^*\right) &= \frac{\partial f_3}{\partial e_2} + \frac{\partial f_3}{\partial c_2}\frac{\partial c_2^*}{\partial e_2} + \frac{\partial f_3}{\partial s_2}\frac{\partial s_2^*}{\partial e_2} \\
&= \frac{\partial f_3}{\partial e_2} + \frac{\partial f_3}{\partial c_2}\left(w_2\left(1 - s_2^*\right) - w_2 e_2 \frac{\partial s_2^*}{\partial e_2}\right) + \frac{\partial f_3}{\partial s_2}\frac{\partial s_2^*}{\partial e_2} \\
&= \frac{\partial f_3}{\partial e_2} + \frac{\partial f_3}{\partial c_2} w_2 \left(1 - s_2^*\right) + \Theta \frac{\partial s_2^*}{\partial e_2}
\end{aligned}$$

Example

To determine the effect of school-aged wages on schooling, we can take the derivative of the first order condition with respect to w_2:

$$\frac{d}{dw_2}\left(w_2 e_2\right) = \frac{d}{dw_2}\left(\beta \frac{\partial f_3}{\partial s_2}\right)$$

$$e_2 = \beta \frac{\partial^2 f_3}{\partial s_2^2}\frac{\partial s_2^*}{\partial w_2}$$

$$\frac{\partial s_2^*}{\partial w_2} = \frac{e_2}{\beta \frac{\partial^2 f_3}{\partial s_2^2}}$$

And since $e_2, \beta > 0$ and $\frac{\partial^2 f_3}{\partial s_2^2} < 0$ by assumption,

$$\frac{\partial s_2^*}{\partial w_2} = \frac{e_2}{\beta \frac{\partial^2 f_3}{\partial s_2^2}} < 0$$

And likewise, to determine the impact of period 2 human capital on schooling:

$$\frac{d}{de_2}(w_2 e_2) = \frac{d}{de_2}\left(\beta \frac{\partial f_3}{\partial s_2}\right)$$

$$w_2 = \beta \left(\frac{\partial^2 f_3}{\partial s_2 \partial e_2} + \frac{\partial^2 f_3}{\partial s_2^2} \frac{\partial s_2^*}{\partial e_2} \right)$$

$$\frac{\partial s_2^*}{\partial e_2} = \frac{w_2 - \beta \frac{\partial^2 f_3}{\partial s_2 \partial e_2}}{\beta \frac{\partial^2 f_3}{\partial s_2^2}}$$

Since $\frac{\partial^2 f_3}{\partial s_2^2} < 0$ by assumption,

$$\frac{\partial s_2^*}{\partial e_2} = \frac{w_2 - \beta \frac{\partial^2 f_3}{\partial s_2 \partial e_2}}{\beta \frac{\partial^2 f_3}{\partial s_2^2}} > 0 \iff \beta \frac{\partial^2 f_3}{\partial s_2 \partial e_2} > w_2$$

B.2 Rainfall Data Appendix

The rain data used in this paper comes from the University of Delaware, and is interpolated to a .5 by .5 grid across the globe. We use rainfall data from 1976-2008 in our analysis. More information on this data set, and the interpolation algorithms and station data used can be found at `http://climate.geog.udel.edu/~climate/html_pages/download.html#P2009`.

In this appendix we provide additional information and summary statistics for the rainfall data used in the paper.

B.2.1 Construction of the Rain Data

The data was constructed using monthly measures of total precipitation from a dense network of rainfall stations across India. The data was interpolated using “Climatologically Aided Interpolation” (CAI), which uses climatological models to help interpret, smooth, and extrapolate data received from rainfall stations. The interpolation is done using Shepard’s algorithm, which employs an enhanced distance weighting method. The number of nearby stations used to interpolate a grid point's precipitation is 20. Cross-validation exercises are performed in which stations are removed one by one, and their precipitation is interpolated using the above methodology. The errors are small due to both the density of stations and the large number of stations used for interpolation.

B.2.2 Construction of the Shock Variables

Following the literature on the effect of rainfall on wages in India, (Jacoby and Skoufias (1997); Kochar (1999); Rose (2001); Jayachandran (2006); Kaur (2011)), we construct a relative measure of "rain shocks". We code years which are above the 80th percentile of rainfall in a given district as "positive shock" years and years which are below the 20th percentile as "negative shock" years. Thus, by construction, each district has 20% positive shock years, 20% negative shock years, and 60% regular years. This construction fits both the wage and crop productivity data well, as is evidenced in Jayachandran (2006) and Kaur (2011).

There are a few things to note about this type of "shock" variable. First, it is by construction relative. That means our comparisons will not be across consistently arid regions (like Rajasthan) and balmier ones (like Kerala), but rather across relatively good and bad years for a given district. We think the relative measure is the right one for two reasons. First, rural agricultural practices are often adapted over many years for the given climate. So if Keralan farmers are growing rice which needs a large amount of rainfall, then it is not reasonable to compare their rainfall levels to those of Rajasthani farmers who are growing different crops which need less water. In addition, we want to capture responses to abnormal, or unexpected events. If the median rainfall in a given area is a "drought" condition, then it's not clear what behavior change we should expect at this level of rainfall.

The other thing to note about our definition of rainfall shocks is that these are not large deviations. By construction, on average one in every five years is a "drought". Thus, when interpreting these results, one should not think about the effects a once-in-a-lifetime crop-destroying event, but rather just a bad or good year. As we show in our empirical analysis, and as has been shown in earlier work, these deviations are still enough to substantially move wages and crop productivity. But particularly when interpreting the magnitude of our early life results, it is important to understand that these "droughts" are not the Dutch famine or the Spanish flu (Almond (2006)). They are simply times of relative scarcity faced at some point during childhood for nearly all rural Indian children.

B.2.3 Summary Statistics

In Table 1, we show summary statistics for total rainfall, rainfall quintile, and total negative and positive shocks from 1989-2008, which roughly corresponds to the time period studied in the ASER data. For the first six variables, an observation is a district-year, and for the last two, an observation is a district. Raw rainfall numbers vary quite a bit over this sample which is not surprising given the size of India and the length of time considered. The quartile variables are all relative to their own district rainfall histories, which accounts for their uniformity. It is not surprising that they hover around .2 since quintile variables are constructed from this data over a longer time horizon. Still, it is reassuring to see that rainfall does not appear to be dramatically increasing or decreasing over this time span. In addition, we report the total number of shocks over this 20-year period for each district. The average for each positive and negative shocks is about 3.5. That is, over the 20-year period, about a third of the time was either a "positive shock" or a "negative shock" and most children will have experienced at least one of each shock during their childhood.

B.2.4 Correlation over Time

One possible issue with using negative and positive rainfall as quasi-random shocks is that they may be correlated over time. There are certainly districts in which droughts are more common in all years, but this should not affect our empirical results since the district fixed effects model uses within district variation in timing of droughts to identify causal effects. However, if it is the case that droughts this year are correlated with droughts next year, then it is difficult to tell the extent to which we are picking up the effects of a single shock or multiple years of rainfall shocks. We test for serial correlation directly in Table 3. In column 1 we find no significant evidence of serial correlation. In column 2 once we include year fixed effects, the coefficient becomes negative and statistically significant, however, the magnitude of the effect is very small. It is unlikely that such a small amount of negative rainfall correlation will affect our results particularly because it means that children exposed to a drought this year are *less* likely to have been exposed to a drought last year. Still, we include an indicator for rainfall shock last year in all regressions.

B.2.5 Spatial Correlation

Our outcome data is at the level of the district. There are 687 relatively uniformly sized districts in India. The median land area of a district is 3,800 square kilometers. On average, a district in our data contains four rainfall grid point measurements. For the analysis of the paper, we use the rainfall measurement at the grid point that is the closest to the geographic center of the district. If there is significant within-district variation in rainfall, our district-level measure of rainfall variation might be missing the true effects for many of the children in our sample. This would attenuate our results, since we would be mis-categorizing the true rainfall for some children. In addition, measurement error at the station level could also lead to attenuation. Since we do not have the original, station-level rainfall data, we cannot investigate either of these phenomena directly, but we can shed some light on the extent of spatial variation in our rainfall data.

In Tables 4-8, we report the correlations between each district's nearest grid points rainfall and the seven next closest grid points rainfall. The correlation between the two nearest grid points is .803, which is reasonably high (for comparison, the correlation between the same grid points in January and February is .38). In Tables 5-8, we report this same statistic for positive shocks, negative shocks, rainfall quintile, and "rain shock". Overall, though the spatial correlation of shocks is high, particularly for the three nearest points, it is unlikely that our results are being overly attenuated by very local variation. This makes intuitive sense—while we are all aware of the phenomenon of local variation in weather patterns on a given day, overall levels of rain in a year tend to be driven by larger weather

Table B.1: *Rainfall Summary Statistics 1989-2008*

	Mean	SD	Min	Max	Percentiles			Obs.
					25	50	75	
Rainfall (mm)	1291	808	34	6455	757	1080	4085	19,618
Rain Quartile 1	.18	.38	0	1	0	0	0	19,618
Rain Quartile 2	.21	.40	0	1	0	0	0	19,618
Rain Quartile 3	.21	.40	0	1	0	0	0	19,618
Rain Quartile 4	.21	.40	0	1	0	0	0	19,618
Rain Quartile 5	.21	.40	0	1	0	0	0	19,618
Total Negative Shocks	3.39	1.27	0	6	3	3	4	577
Total Positive Shocks	3.47	1.35	0	7	3	3	4	577

patterns that persist over larger areas.

B.3 ASER Data Appendix

In 2005, Pratham, an Indian NGO, initiated the Annual Status of Education Report (ASER), a research program unlike any ever undertaken in India. Pratham's ASER, is India's largest privately-funded survey. Every year ASER, meaning "impact" in Hindi, measures the enrollment as well as the reading and arithmetic levels of Indian children ages 5 to 16 in every rural district of India. The purpose of ASER is twofold: (i) to get reliable estimates of the status of children's schooling and basic learning (reading, writing and math ability) at the district level; and (ii) to measure the change in these basic learning and school statistics from year to year.

B.3.1 Test Scores

Every child ages 5-16 in a household is tested. All children are given the same math and reading questions described below. The test is designed to measure core competencies that a child should have mastered by standards 1-2 (grades 1-2). Vagh (2009) presents evidence for the psychometric properties of the ASER testing tools, and shows that the reading tests are highly correlated with other assessment tests.

All children regardless of whether they are enrolled in school are tested at home. Enumerators are instructed to visit a random sample of households only when children are

Table B.2: *Percent of Droughts and Positive Rainfall Shocks by Year*

Year	% Top Quartile Rainfall	% Bottom Quartile Rainfall
1975	.35	.03
1976	.16	.17
1977	.29	.09
1978	.29	.14
1979	.03	.46
1980	.13	.22
1981	.11	.15
1982	.06	.30
1983	.26	.08
1984	.26	.17
1985	.26	.16
1986	.12	.26
1987	.24	.35
1988	.44	.05
1989	.13	.15
1990	.43	.02
1991	.11	.19
1992	.01	.45
1993	.14	.15
1994	.29	.05
1995	.11	.13
1996	.11	.19
1997	.12	.15
1998	.20	.03
1999	.07	.22
2000	.03	.22
2001	.04	.14
2002	.02	.42
2003	.08	.14
2004	.06	.24
2005	.19	.17
2006	.20	.30
2007	.25	.04
2008	.29	.05

Notes: This table shows estimates of the percent of districts each year that experience a drought and positive rainfall shock.

Table B.3: *Testing for Serial Correlation in Rainfall 1988-2008*

Dependent Variable: Deviation from district mean this year	(1)	(2)
Deviation from district mean last year	.005	-.031***
	(.011)	(.010)
Year Fixed Effects	N	Y
Observations	9,248	9,248

Notes: This table tests if there is serial correlation in rainfall in our data. An observation is a district year. The dependent variable in both regressions is the deviation from mean rainfall in the current year (in inches), where deviation is simply defined as current year rainfall minus the mean rainfall in sample period. The independent variable is deviation from mean rainfall last year (in inches), constructed in the same way. The mean of the deviation is 0 (2.2e-06)and the standard deviation is 223 inches. Standard errors are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.4: *Spatial Correlation: Raw Rainfall*

	Closest Grid Points:							
	1	2	3	4	5	6	7	8
1	1							
2	.803	1						
3	.827	.680	1					
4	.805	.763	.742	1				
5	.787	.664	.634	.616	1			
6	.665	.783	.519	.579	.607	1		
7	.847	.715	.749	.736	.689	.588	1	
8	.812	.645	.774	.724	.692	.565	.745	1

Notes: This table shows the correlation between the raw rainfall grid measurements in mm nearest to the center of the district. (1) represents the closest grid point to the center, (2) represents the second closest, and so on. Data is from 1998-2008.

Table B.5: *Spatial Correlation: Negative Shock*

	Closest Grid Points:							
	1	2	3	4	5	6	7	8
1	1							
2	.742	1						
3	.749	.664	1					
4	.682	.713	.709	1				
5	.732	.666	.653	.580	1			
6	.673	.717	.606	.601	.678	1		
7	.694	.637	.641	.610	.656	.580	1	
8	.812	.645	.774	.724	.692	.565	.745	1

Notes: This table shows the correlation between the measure of negative shocks at the grid points nearest to the center of the district. (1) represents the closest grid point to the center, (2) represents the second closest, and so on. Data is from 1998-2008.

Table B.6: *Spatial Correlation: Positive Shock*

	Closest Grid Points:							
	1	2	3	4	5	6	7	8
1	1							
2	.760	1						
3	.754	..689	1					
4	.706	.734	.729	1				
5	.733	.671	.655	.598	1			
6	.693	.724	.635	.618	.684	1		
7	.695	.652	.645	.630	.641	.603	1	
8	.661	.591	.722	.612	.601	.570	.641	1

Notes: This table shows the correlation between the measure of positive shocks at the grid points nearest to the center of the district. (1) represents the closest grid point to the center, (2) represents the second closest, and so on. Data is from 1998-2008.

Table B.7: *Spatial Correlation: Rainfall Quintile*

	Closest Grid Points:							
	1	2	3	4	5	6	7	8
1	1							
2	.894	1						
3	.897	.853	1					
4	.857	.874	.870	1				
5	.881	.839	.825	.774	1			
6	.855	.870	.802	.795	.845	1		
7	.862	.816	.825	.798	.824	.782	1	
8	.840	.789	.870	.788	.791	.759	.813	1

Notes: This table shows the correlation between the quintile of rainfall at the grid points nearest to the center of the district. (1) represents the closest grid point to the center, (2) represents the second closest, and so on. Data is from 1998-2008.

Table B.8: *Spatial Correlation: Rain shock*

	Closest Grid Points:							
	1	2	3	4	5	6	7	8
1	1							
2	.815	1						
3	.816	.757	1					
4	.774	.795	.790	1				
5	.799	.748	.734	.684	1			
6	.758	.786	.705	.699	.753	1		
7	.768	.721	.726	.705	.727	.679	1	
8	.742	.687	.780	.692	.691	.657	.721	1

Notes: This table shows the correlation between rain shock at the grid points nearest to the center of the district. (1) represents the closest grid point to the center, (2) represents the second closest, and so on. Data is from 1998-2008.

likely to be at home; they must go on Sundays when children are not in school and no one works. If all children are not home on the first visit, enumerators are instructed to revisit once they are done surveying the other households (ASER, 2010). We have test score data from 2005-2009.

Reading task

All children are assessed using a simple reading tool.[1] Similar tests are developed in all languages each year. There are approximately 16 distinct regional languages. The child can choose the language in which she wants to read (Hindi vs. local language). In developing these tools in each state language, care is taken to ensure comparability with the previous years' tool with respect to word count, sentence count, type of word and conjoint letters in words, compatibility with the vocabulary and sentence construction used in Standard 1 and Standard 2 language textbooks of the state, and familiarity with words and context through extensive field piloting. The reading test has 4 questions:

1. Letters: Child is asked to identify a set of common letters. The child is asked to read any 5 letters from a letters list. The child should choose the letters herself. If she does not choose, then the enumerator should point out letters to her. If she can correctly recognize at least 4 out of 5 letters with ease, then she is marked as a child who "can read letters". If she cannot read 4 out of 5 letters correctly, then she is marked as a child who "cannot recognize letters".

2. Words: The child is asked to read 5 words from a word list. These are 1 or 2 matra (beats or syllables) words. The child should choose the words herself. If she does not choose, then the enumerator points out words to her. If she can correctly read at least 4 out of the 5 words with ease, then she is marked as mastering the "word" level.

3. Level 1 (Standard 1) text: Child is asked to read a set of 4 simple linked sentences. Each sentence should have no more than 4-5 words. These words or their equivalent

[1] A sample reading and math exam from 2009 ASER are attached at the end of the Data Appendix.

are in the Standard 1 textbook of the state. The child may read slowly. She may read haltingly; she may make 3 or 4 mistakes in not reading words correctly. However, as long as the child reads the text like she is reading a sentence, rather than a string of words, she should be marked as a child who "can read Level 1 text". If the child stops very often, has difficulty with more than 3 or 4 words and reads like she is reading a string of words not a sentence, then she is marked as a child who cannot read Level 1 text.

4. Level 2 (Standard 2) text: Child is asked to read a "short" story with 7-10 sentences. Sentence construction is straightforward, words are common and the context is familiar. These words (or their equivalent) are in the Standard 2 textbook of the state. If she can read fluently with ease, then mark her as a child who "can read Level 2 text".

Math task

All children are also assessed using a simple arithmetic tool. The arithmetic test has 4 questions:

1. Number recognition 1 to 9: Child is asked to identify randomly chosen numbers from 1 to 9. If child can correctly identify at least 4 out of 5 numbers then she is marked as a child who can "recognize numbers from 1-9." If not, she is marked as a child who "cannot recognize numbers".

2. Number recognition 11 to 99: Child is asked to identify randomly chosen numbers from 11 to 99. If child can correctly identify at least 4 out of 5 numbers then she is marked as a child who can "recognize numbers from 11-100."

3. Subtraction: Child is asked to complete 2 digit numerical problems with borrowing. If child can complete two subtraction problems correctly, she is asked to attempt the division problem.

4. Division: Child is asked to complete a 3 digit by 1 digit numerical division problem. If she can complete one, then she is marked as knowing how to do division. If she

cannot complete the first one, she is asked to try another. If she cannot do the second one correctly either, then she is marked as being unable to do division.

Math word problem

In 2006 and 2007, children were also asked to complete two math word problems in addition to the math and reading tasks described above. Like the reading tasks, similar math tests are developed in all languages. The two math word questions in the 2007 ASER were (ASER, 2007):

1. You have Rs. 50. From that you buy a pair of shoes for Rs.35. How much money do you have left with you now?
2. You have Rs. 50. From that you buy sweets for Rs. 28. How much money do you have left with you now?

For each measure of math and reading, children were marked as being able to complete the task or not (i.e. they get a 0 if they cannot complete the task and a 1 if they complete correctly). For each of our measures, we simply add up the number of questions the child answered correctly and assign that score. For example, a child who gets all 4 math questions correct, receives a 4, and a child who gets none of them right, receives a 0. For the math word problems, the maximum score is a 2 and for the reading score it is a 4. We do note that the reading and math tasks get progressively harder from question 1 to 4. Therefore, simply adding them up as opposed to using an ordinal scale might be problematic. Therefore, we also estimate the regressions using ordered logit models, and the results are similar (results are in appendix).

One issue with the math tasks is that in 2005 and 2006 students were only asked to complete questions 2, 3 and 4. They were not asked the simple number recognition problem. Therefore in 2005 and 2006, we assign the children who were unable to complete questions 2-4 a 0.65 as they would have either been categorized as a 0 or 1. This amount, 0.65, is the number of children who got question 1 correct divided by the total number of children who scored a 0 or 1 in the other years. We also estimate the regressions leaving out the 2005 and 2006 children in case there is concern that this adjustment impacts the results.

We also note that there is quite a bit of variation across ages and questions. One might be concerned that all of the 5 year olds are getting all of the questions wrong and all of the 16 year olds are getting all of the questions correct. This is not the case. In Figure 1 we plot the math (top panel) and reading scores (bottom panel) by age where young is ages 5-8, middle is ages 9-12, and old is ages 13-16. Figure 1 shows that children ages 13-16 constitute 10 percent of the children scoring 0 on the math and reading tests and only constitute 50 percent of the highest scores for both math and language. Therefore, there is quite a lot of variation in scores across all ages.

B.3.2 Sampling

The sampling strategy used generates a representative picture of each district. All rural districts in India (slightly less than 600) are surveyed. The survey is designed to be a household survey. Within each district, 30 villages are randomly chosen and in each village 20 households are randomly picked for a total of 600 households per district. A rotating panel of villages (rather than children) is adopted. Each year, 10 villages from 2 years ago are dropped and 10 new villages added. For instance, in ASER 2009 the 10 villages from ASER 2006 were dropped, the 10 villages from 2007 and 2008 were retained and 10 new villages from the census village directory of 2001 were added (ASER, 2013).[2] A sample size of 600 households gives us approximately 1000–1200 children per district.

The 600 households could be randomly selected if household lists at the district level exist. However, in the absence of these, a two-stage sample design was adopted. In the first stage, 30 villages were randomly selected using the village directory of the 2001 census as the sample frame. In the second stage 20 households were randomly selected in each of the 30 selected villages in the first stage (ASER, 2013).

Villages were selected using the probability proportional to size (PPS) sampling method.

[2]Villages are chosen from the 2001 Census Directory using PPS (Probability Proportional to Size).

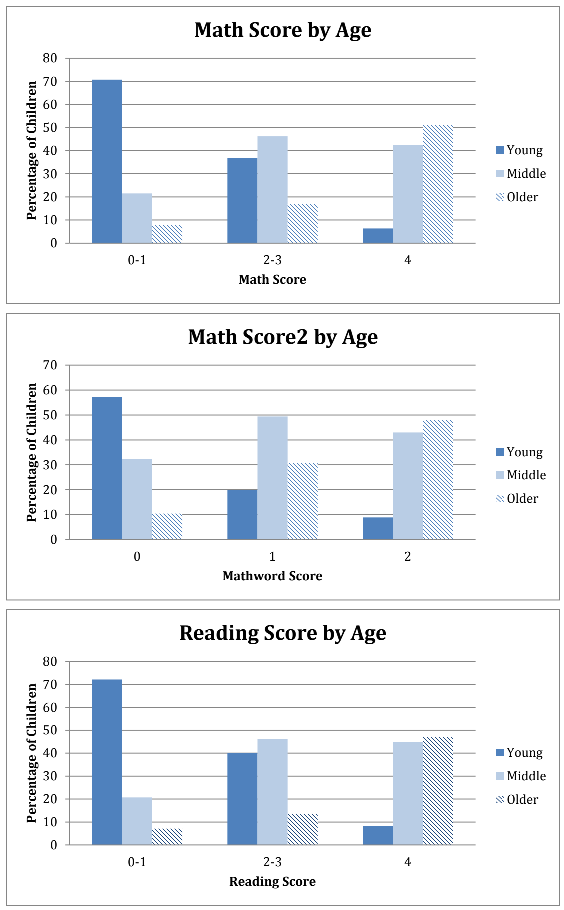


Figure B.1: *Math and Reading Test Scores by Age*

This method allows villages with larger populations to have a higher chance of being selected in the sample. It is most useful when the sampling units vary considerably in size because it assures that those in larger sites have the same probability of getting into the sample as those in smaller sites, and vice versa.[3]

In the selected villages, 20 households are surveyed. Ideally, a complete list of all households in the selected village should have been made and 20 households selected randomly from it. However, given time and resource constraints a procedure for selecting households was adopted that preserved randomness as much as possible. The field investigators were asked to divide the village into four parts. This was done because villages often consist of hamlets and a procedure that randomly selects households from some central location may miss out households on the periphery of the village. In each of the four parts, investigators were asked to start at a central location and pick every 5th household in a circular fashion until 5 households were selected.[4]

The key feature of ASER is that it is a rapid assessment survey. As a result, the survey instrument is short and its focus is on the assessment of basic learning. Since 2006, ASER has included younger children in the sample. However, children ages three and four are not tested. All that they are asked is whether they attend any kind of pre-school (such as anganwadi). Older children (aged 5 to 16) are queried about school enrollment and are tested in basic reading and arithmetic. These tests have been carried through all the ASER surveys, and this is the data we use in our study.

B.4 Appendix Tables

[3] Most large household surveys in India, like the National Sample Survey and the National Family Health Survey also use this two stage design and use PPS to select villages in the first stage.

[4] In larger villages, the investigators increased the interval according to a rough estimate of the number of households in each part.

Table B.9: *Effect of Rainfall Shocks on Math Scores: Full and Restricted Sample*

	Dependent Variable:			
	Math Score (Full Sample)	Math Score (Restricted)	Math Score (Full Sample)	Math Score (Restricted)
Rain Shock This Year	-.02 (.01)*	-.03 (.02)		
Rain Shock Last Year	-.02 (.01)*	-.04 (.02)**		
Rain Shock In Utero			.01 (.004)***	.003 (.004)
Rain Shock Year of Birth			.01 (.004)***	.005 (.004)
Rain Shock at Age 1			.01 (.004)***	.01 (.004)**
Rain Shock at Age 2			.01 (.004)**	.01 (.004)***
Rain Shock at Age 3			.001 (.004)	.01 (.004)**
Rain Shock at Age 4			.002 (.004)	.01 (.004)***
Observations	2,109,162	1,194,128	2,356,028	1,748,743

Notes: This table shows our estimates of the effect of rainfall shocks on math test scores. Column 1 is the same as column 1 in Table 3 and column 3 is the same as column 1 in Table 6. In columns 2 and 4, we restrict the samples to include post–2006 ASER data. All columns include year and age fixed effects. Columns 1-2 also include district fixed effects and columns 3-4 also contain household fixed effects. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.10: *Drought and Crop Yields: 1957-1987*

	Dependent Variable:					
	Rice		Wheat		Jowar	
Rain Shock	.06 (.02)**	.08 (.02)***	.002 (.01)	.05 (.008)***	.01 (.01)	.02 (.009)***
Year fixed effects	Y	Y	Y	Y	Y	Y
District fixed effects	Y	Y	Y	Y	Y	Y
Controls	Y	N	Y	N	Y	N
Observations	7161	8401	6680	8401	6265	7409
Mean Dependent Variable	1.51	1.51	.856	.856	.589	.589

Source: Data on crop yields and inputs is from World Bank India Agriculture and Climate Data set which has agricultural yield (revenues per acre) data from 1975-1987.. Rainfall data from University of Delaware. Notes: This table shows results from a regression of crop yields on rain shocks. An observation is a district-year. Controls are measures of inputs used in production: labor, bullocks, fertilizer, and machinery, as well as 3-year average yield. Standard errors are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.11: *Percent of Droughts and Positive Rainfall Shocks by Year*

Year	% Top Quartile Rainfall	% Bottom Quartile Rainfall
1975	.35	.03
1976	.16	.17
1977	.29	.09
1978	.29	.14
1979	.03	.46
1980	.13	.22
1981	.11	.15
1982	.06	.30
1983	.26	.08
1984	.26	.17
1985	.26	.16
1986	.12	.26
1987	.24	.35
1988	.44	.05
1989	.13	.15
1990	.43	.02
1991	.11	.19
1992	.01	.45
1993	.14	.15
1994	.29	.05
1995	.11	.13
1996	.11	.19
1997	.12	.15
1998	.20	.03
1999	.07	.22
2000	.03	.22
2001	.04	.14
2002	.02	.42
2003	.08	.14
2004	.06	.24
2005	.19	.17
2006	.20	.30
2007	.25	.04
2008	.29	.05

Source: Rainfall data from the University of Delaware. Notes: This table shows the percent of districts each year that experience a drought and positive rainfall shock.

Table B.12: *Effect of Contemporaneous Rainfall Shocks on Human Capital (Ordered Logit)*

	Dependent Variable:		
	Math Score	Math Word Problem	Read Score
Rain Shock This Year	-.03	-.15	-.002
	(.02)	(.07)**	(.02)
Rain Shock Last Year	-.03	-.15	-.04
	(.02)	(.08)**	(.02)*
Observations	2,109,162	843,827	2,120,708

Source: Test Score Data from ASER 2005-2009. Rainfall Data from University of Delaware. Notes: This table shows ordered logit estimates of the effect of rainfall shocks on current test scores using the ASER data for all children ages 5-16 (or β_1 and β_2 from equation (1)). "Math Score" and "Read Score" range from 0-4. "Math Word Problem" ranges from 0-2 and is only available in 2006 and 2007. All regressions contain fixed effects for district, year and age. All columns contain controls for early life rainfall shock exposure (in utero-age 4). Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.13: *Effect of Contemporaneous Rainfall Shocks on Human Capital (Boys)*

	Dependent Variable:				
	Math Score	Math Word Problem	Read Score	Dropped Out	Attendance
Rain Shock This Year	-.01	-.05	.003	-.0004	-.02
	(.01)	(.02)***	(.01)	(.0008)	(.006)***
Rain Shock Last Year	-.01	-.04	-.02	.002	-.03
	(.01)	(.02)**	(.01)*	(.0009)**	(.008)***
Observations	1145,955	465,547	1152,131	1192,358	250,178
Mean Dependent Variable	2.66	1.281	2.723	0.035	0.863

Source: Test Score Data from ASER 2005-2008. Rainfall Data from University of Delaware. Notes: This table shows estimates of β_1 and β_2 from equation (1), the effect of rainfall shocks on current test scores for boys ages 5-16. Columns 1-4 contain fixed effects for district, year and age. Since attendance is only observed in 2008, column 5 contains fixed effects for state, year, and age. All columns contain controls for early life rainfall shock exposure (in utero-age 4). Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.14: *Effect of Contemporaneous Rainfall Shocks on Human Capital (Girls)*

	Dependent Variable:				
	Math Score	Math Word Problem	Read Score	Dropped Out	Attendance
Rain Shock This Year	-.02	-.05	.002	.0008	-.02
	(.01)*	(.02)***	(.01)	(.0008)	(.006)***
Rain Shock Last Year	-.02	-.04	-.02	.002	-.04
	(.01)*	(.02)*	(.01)*	(.0009)**	(.009)***
Observations	951,233	378,280	956,529	988,483	208,602
Mean Dependent Variables	2.567	1.225	2.662	0.039	0.863

Source: Test Score Data from ASER 2005-2008. Rainfall Data from University of Delaware. Notes: This table shows estimates of β_1 and β_2 from equation (1), the effect of rainfall shocks on current test scores for girls ages 5-16. Columns 1-4 contain fixed effects for district, year and age. Since attendance is only observed in 2008, column 5 contains fixed effects for state, year, and age. All columns contain controls for early life rainfall shock exposure (in utero-age 4). Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.15: *Effect of Early Life Rainfall Shocks on Human Capital (Boys)*

	Dependent Variable:				
	Math Score	Math Word Problem	Read Score	Never Enrolled	On Track
Rain Shock In Utero	.008	.003	.01	-.001	.02
	(.004)*	(.006)	(.005)**	(.0004)***	(.002)***
Rain Shock Year of Birth	.01	.01	.008	-.002	.02
	(.004)**	(.006)*	(.004)*	(.0004)***	(.002)***
Rain Shock at Age 1	.01	.02	.01	-.002	.01
	(.005)**	(.006)***	(.005)**	(.0005)***	(.002)***
Rain Shock at Age 2	.01	.02	.01	-.003	.01
	(.004)***	(.006)***	(.004)***	(.0004)***	(.002)***
Rain Shock at Age 3	.006	.01	.01	-.002	-.0002
	(.005)	(.006)**	(.005)**	(.0004)***	(.002)
Rain Shock at Age 4	.003	-.008	.01	-.002	.002
	(.005)	(.006)	(.005)**	(.0004)***	(.002)
Observations	1,271,233	465,547	1,277,571	1,297,538	959,304
Mean Dependent Variable	2.66	1.281	2.723	0.026	0.811

Source: Test Score Data from ASER 2005-2009. Rainfall Data from University of Delaware. Notes: This table shows estimates of ζ from equation (2), the effect of early life rainfall shocks on current test scores and schooling outcomes restricted to boys only. “Math Score” and “Read Score” range from 0-4. “Math Word Problem” ranges from 0-2, and was only asked in 2006 and 2007. “On Track” is equal to one if age minus grade is at least six, and zero otherwise. All regressions contain fixed effects for household, year and age. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.16: *Effect of Early Life Rainfall Shocks on Human Capital (Girls)*

	Dependent Variable:				
	maths score	Math Word Problem	Read Score	Never Enrolled	On Track
Rain Shock in Utero	.02	.009	.02	-.002	.03
	(.005)***	(.006)	(.005)***	(.0006)***	(.003)***
Rain Shock Year of Birth	.01	.005	.01	-.002	.02
	(.005)**	(.007)	(.006)**	(.0006)***	(.003)***
Rain Shock at Age 1	.01	.01	.02	-.003	.02
	(.005)***	(.007)	(.005)***	(.0005)***	(.003)***
Rain Shock at Age 2	.007	.01	.01	-.003	.01
	(.005)	(.007)**	(.005)**	(.0006)***	(.003)***
Rain Shock at Age 3	-.005	-.002	.002	-.001	-.0002
	(.005)	(.007)	(.005)	(.0006)*	(.002)
Rain Shock at Age 4	-.003	-.01	.009	-.002	.0000306
	(.005)	(.007)*	(.005)*	(.0005)***	(.002)
Observations	1,057,467	378,280	1,062,888	1,079,939	808,469
Mean Dependent Variable	2.567	1.225	2.662	0.035	0.811

Source: Test Score Data from ASER 2005-2009. Rainfall Data from University of Delaware. Notes: This table shows estimates of ζ from equation (2), the effect of early life rainfall shocks on current test scores and schooling outcomes restricted to girls only. “Math Score” and “Read Score” range from 0-4. “Math Word Problem” ranges from 0-2, and was only asked in 2006 and 2007. “On Track” is equal to one if age minus grade is at least six, and zero otherwise. All regressions contain fixed effects for household, year and age. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.17: *Effect of Rain Shocks on Days Sick and Health Expenditures*

	Days sick	Health expenditures (ln Rs)
Rain Shock This Year	-.52	-.22
	(.22)**	(.07)***
Rain Shock Last Year	-.38	.005
	(.22)*	(.1)
Observations	6293	6293
Mean Dependent Variable	6.07	4.28

Source: IHDS 2004–2005 data for children ages 5-16. Rainfall data from the University of Delaware. Notes: This table shows estimates of the effect of rainfall shocks on number of days sick in last month due to diarrhea, fever, and/or cough and health expenditures (hospital, doctor, medicine, tests, and transport) for children ages 5-16. Each cell is a separate OLS regression. All regressions contain age, gender and state fixed effects. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.18: *Effect of Rain Shocks on Test Scores in High Malaria States*

	Dependent Variable: Math Score	Reading Score
Rain shock	-.08	-.03
	(.03)**	(.03)
Malaria state	-.14	-.1
	(.12)	(.12)
Rain shock*Malaria state	.07	.03
	(.07)	(.06)
Rain Shock Last Year	.04	.02
	(.03)	(.03)
Observations	1,892,741	2,115,547

Source: Test Score Data from ASER 2005-2009. Rainfall Data from University of Delaware. Notes: This table shows estimates β_1, the effect of rainfall shocks on current test scores and schooling outcomes interacted with a dummy for the five high malaria states (Orissa, Chhattisgarh, West Bengal, Jharkhand, and Karnataka). All specifications include state-region fixed effects and are clustered at the state level. All columns contain controls for early life rainfall shock exposure (in utero-age 4). Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.19: *Are Teacher Absences or School Lunches Driving the Results?*

Dependent Variable:	Teacher Absence Rate	Midday Meal Provision
Rain Shock	-.03	.04
	(.01)**	(.02)**
Rain Shock Last Year	.002	.06
	(.01)	(.02)***
Observations	20,297	24,203
Mean Dependent Variable	0.18	0.81

Source: Teacher absence rates and midday meal provision data from the 2005 and 2007 ASER School Survey. Notes: This table shows the coefficients of rainfall shocks on teacher absence rates and midday meal provision in a linear regression. All regressions contain village and year fixed effects. Standard errors, clustered at the district level, are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.20: *Does Drought Impact Fertility Decisions?*

	ln cohort size (born 1991) (1)	ln cohort size (born 2001) (2)	ln cohort size (born 2001) (3)
Drought (t)	-.01	-.01	-.01
	(.02)	(.02)	(.02)
Drought In utero (t-1)	.04	-.002	-.005
	(.06)	(.03)	(.03)
Drought (t-2)	.008	-.04	-.04
	(.02)	(.02)	(.02)
Drought (t-3)	-.04	.03	.03
	(.04)	(.06)	(.06)
Drought (t-4)	-.02	.09	.09
	(.02)	(.03)***	(.03)***
Drought (t-5)	-.03	-.02	-.02
	(.03)	(.03)	(.03)
ln Population 1991	.04		
	(.02)**		
ln Population 2001		.02	
		(.02)	
ln Female Population 2001 (15-49)			.01
			(.02)
Observations	104,631	207,905	205,728
Mean Dependent Variable	5.33	5.98	5.98

Source: Number of births in each district in 1991 and 2001 using Indian Census Data. Rainfall data from the University of Delaware. Notes: These are OLS regressions where the dependent variable is ln number of births in each district in 1991 and 2001. All regressions contain state and year of survey fixed effects. Standard errors are clustered at the district level and are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table B.21: *Effect of Rain Shocks on Migration Rates*

	Has Not Moved (Last Six Months)			
	Full Sample	Ages 5-16	Full Sample	Ages 5-16
Drought This Year	.001	.002		
	(.003)	(.003)		
Drought Last Year	.006	.006		
	(.003)*	(.004)		
Positive Shock This Year			-.003	.007
			(.01)	(.009)
Positive Shock Last Year			.0002	-.001
			(.007)	(.007)
Observations	236,429	67,521	236,429	67,521
Mean Dependent Variable	.987	.987	.987	.987

Source: Migration data from NSS round 55 (1999-2000). Rainfall data from the University of Delaware. Notes: These are OLS regressions where the dependent variable is has not moved from district in past six months or more, and the independent variable is rain shocks. In odd numbered columns we use the entire sample and in all even numbered columns we restrict the sample to children ages 5-16. All regressions contain state fixed effects. Standard errors are clustered at the district level and are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Appendix C

Appendix to Chapter 3

C.1 Appendix Tables

Table C.1: *Robustness: Effect of NREGA on Test Scores and Schooling*

Dep. Var:	Math Score	Math Score2	Reading Score	Currently Enrolled	Current Grade
			Panel A: Full Sample		
NREGA	-.006	-.08	.005	-.003	-.1
	(.02)	(.02)***	(.02)	(.003)	(.02)***
Observations	2,666,336	2,666,336	2,681,363	2,768,347	2,581,262
			Panel B: Ages 13-16		
NREGA	-.01	-.09	-.02	-.003	-.09
	(.02)	(.02)***	(.02)	(.004)	(.04)**
Observations	720,903	720,903	723,264	765,647	677,122
Mean DV	2.48	3.17	3.57	.87	7.66
			Panel C: Ages 5-12		
NREGA	.007	-.05	.02	-.001	-.02
	(.02)	(.02)**	(.02)	(.002)	(.02)
Observations	1,945,433	1,945,433	1,958,099	2,002,700	1,904,140
Mean DV	1.55	1.84	2.40	.96	3.51
Dist FEs	YES	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES	YES
Wave*Year FE	YES	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on math and reading test scores and schooling outcomes from the ASER data. All columns contain data from 2005-2009. Math score ranges from 0-3 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide(=3), Math Score2 ranges from 0-5 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide (=3), Can do first word problem (=4), Can do second word problem 2 (=5), Read score ranges from 0-4 (Nothing(=0), Read Letters(=1), Read Words(=2), Read Paragraph(=3) and Read Story(=4). All Panels include year and district fixed effects and wave*year fixed effects. Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table C.2: *Effect of NREGA on Working and School Attendance (NSS)*

	Ages 13-17			
Years:	2003-2009 (Actual)		1999-2005 (Placebo)	
Dep. Var:	Works	Attends School	Works	Attends School
NREGA	.03	-.03	-.005	-.005
	(.008)***	(.008)***	(.009)	(.009)
Observations	183,915	183,218	115,553	115,553
Mean of DV	.25	.68	.29	.66
	Ages 5-12			
NREGA	.001	-.0002	-.005	-.004
	(.003)	(.007)	(.002)**	(.009)
Observations	331,715	331,715	214,766	214,766
Mean of DV	.03	.82	.03	.79
Dist FEs	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on children working and attending school. Columns 1-2 use data from the NSS 2003-2009 (rounds 60, 61, 62, 64, 66). The placebo analysis utilizes NSS 1999-2005 data (rounds 55, 60, 61, 62) where NREGA equals 1 in 2004. Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table C.3: *Effect of NREGA in High Job Districts on Schooling and Test Scores, Ages 13-17*

Dep. Var:	Math Score	Math Score2	Reading Score	Currently Enrolled	Current Grade
	Panel A: Full Sample				
NREGA	-.01	-.08	-.03	-.13	-.007
	(.02)	(.04)**	(.02)	(.05)**	(.006)
NREGA X High # Jobs	-.01	-.07	-.003	.03	-.002
	(.03)	(.05)	(.03)	(.07)	(.008)
Observations	720,903	721,447	723,264	677,122	765,647
Dist FEs	YES	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA, and NREGA interacted with being a district who provided above median jobs per capita, on math and reading test scores and schooling outcomes from the ASER data for children age 13-16. All columns contain data from 2005-2009. "High # jobs" is equal to one if the district's 2009 number of total job cards divided by its 2011 population is above the median for all districts. Math score ranges from 0-3 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide(=3), Math Score2 ranges from 0-5 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide (=3), Can do first word problem (=4), Can do second word problem 2 (=5), Read score ranges from 0-4 (Nothing(=0), Read Letters(=1), Read Words(=2), Read Paragraph(=3) and Read Story(=4). All Panels include year and district fixed effects. Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table C.4: *Children's Primary Activity (Ages 5-12)*

Dep. Var:	Works at Home	Works outside Home	Domestic Work	Attends School
NREGA	-.002	.0008	.003	-.0002
	(.002)	(.0007)	(.002)	(.007)
Observation	331,715	331,715	331,715	331,715
Mean of DV	.008	.005	.017	.82
Dist FEs	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA . Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table C.5: *Effect of NREGA on Schooling and Time Use*

Dep. Var:	School Time	HH Chore Time	Farm Chore Time	Caring for HH Members	Study Time	Leisure Time	Missing Time
				Panel A: Full Sample			
NREGA	-.02	.01	.002	-.005	.005	.009	.006
	(.008)***	(.003)***	(.004)	(.002)**	(.006)	(.01)	(.007)
Observations	2205	2205	2205	2205	2205	2205	2205
Mean DV	0.27	0.02	.007	.009	.065	.57	.029
				Panel B: Older Cohort			
NREGA	-.03	.02	.007	-.008	-.01	.01	.02
	(.02)**	(.007)***	(.01)	(.005)*	(.009)*	(.02)	(.01)**
Observations	748	748	748	748	748	748	748
Mean DV	.25	.05	.019	.010	.078	.53	.027
				Panel C: Younger Cohort			
NREGA	-.02	.001	-.0008	-.004	.02	.006	-.003
	(.008)*	(.002)	(.0008)	(.003)*	(.007)**	(.01)	(.009)
Observations	1457	1457	1457	1457	1457	1457	1457
Mean DV	.28	.009	.0004	.009	.057	0.62	.03
Child FEs	YES	YES	YES	YES	YES	YES	YES
Round FEs	YES	YES	YES	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on schooling outcomes and time use from the Young Lives data. This is a panel of 2 cohorts of children (young and old): round 1 is from 2002, round 2 from 2006 and round 3 from 2009. The time use questions were asked in rounds 2 and 3 for both cohorts. They are hours/day in a 24 hour period spent in each activity. Leisure includes time spent sleeping and playing. All time use categories should add up to 24 hours. When they do not, the variable missing time was generated to account for the missing hours/day. All regressions include child and round fixed effects. Standard errors clustered at the child level and are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table C.6: *Effect of NREGA on Migration Rates (Age 13-17)*

Dep. Var:	Temporary Migration		
Sample:	All Children	Boys	Girls
NREGA	.01	.02	.003
	(.004)***	(.005)***	(.004)
Observations	73,773	39,333	34,440
Mean DV	.018	.026	.009
District FEs	YES	YES	YES
Year FEs	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on temporary migration rates amongst adolescents in the NSS. We use data from rounds 55 and 64 (these are the only rounds with data on temporary migration). The dependent variable is a binary variable which is equal to one if an individual has migrated for at least one but not more than six months in the past year. All columns contain district and year fixed effects. Standard errors clustered at the district-year level and are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table C.7: *Effect of NREGA on Child Health*

Dep. Var:	Weight-for-age Z-score	Height-for-age Z-score	Had illness Since Visit	Has Long-term Health Problem	Child Same or Better Health
NREGA	.15	.3	-.24	-.05	.07
	(.05)***	(.06)***	(.04)***	(.02)**	(.03)**
Observations	2280	2280	2280	2280	2280
Mean DV					
Child FEs	YES	YES	YES	YES	YES
Round FEs	YES	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on health outcomes using the Young Lives data. This is a panel of 2 cohorts of children (young and old): round 1 is from 2002, round 2 from 2006 and round 3 from 2009. Weight was collected in round 1 for the older cohort and rounds 1-3 for the younger cohort. Height was collected in all rounds for both cohorts. Had illness since visit equals 1 if the child had a serious injury or illness since the last enumerator visit and it was asked of both cohorts in rounds 1 and 2. Has Long-term health problem equals 1 if the child has a long-term health illness or health problem, and it was asked of both cohorts in rounds 1 and 2. Child same or better health equals 1 if the child's health compared to peers is at least the same or better, and it was asked of both cohorts in rounds 1 and 2. All Panels include child and round fixed effects. Standard errors clustered at the child level and are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.

Table C.8: *Effect of NREGA on Test Scores and Schooling Interactions*

Dep. Var:	Math Score	Math Score2	Reading Score	Currently Enrolled	Current Grade
	Panel A: Full Sample, All Ages				
NREGA	.002	.08	.008	-.004	-.02
	(.01)	(.02)***	(.02)	(.003)	(.02)
Oldest girl	.03	.21	.04	-.007	-.03
	(.004)***	(.008)***	(.006)***	(.001)***	(.01)***
NREGA*Oldest girl	-.02	-.24	.0004	-.003	-.01
	(.005)***	(.009)***	(.007)	(.002)	(.01)
Oldest boy in household	.05	.27	.06	.007	.01
	(.005)***	(.009)***	(.006)***	(.001)***	(.009)*
NREGA*Oldest boy	-.02	-.28	-.004	-.007	-.04
	(.006)***	(.01)***	(.008)	(.001)***	(.01)***
# of children in household	-.008	-.01	-.01	-.006	-.03
	(.001)***	(.002)***	(.002)***	(.0003)***	(.002)***
Observations	2,358,118	2,358,118	2,369,967	2,414,033	2,251,174
	Panel B: Full Sample, All Ages				
NREGA	-.007	-.08	.02	-.005	-.01
	(.01)	(.02)***	(.02)	(.003)*	(.02)
Youngest girl	.02	-.13	.04	.01	.07
	(.005)***	(.009)***	(.006)***	(.001)***	(.01)***
NREGA*Youngest girl	.01	.22	-.02	.002	-.01
	(.006)*	(.01)***	(.008)***	(.002)	(.02)
Youngest boy	-.04	-.2	-.02	.002	.01
	(.005)***	(.009)***	(.006)***	(.001)	(.01)
NREGA*Youngest boy	.02	.23	-.01	.001	-.005
	(.006)***	(.01)***	(.008)*	(.002)	(.01)
Oldest girl	.03	.19	.04	-.005	-.03
	(.004)***	(.008)***	(.006)***	(.001)***	(.01)**
NREGA*Oldest girl	-.01	-.22	.002	-.003	-.01
	(.005)***	(.009)***	(.007)	(.002)	(.01)
Oldest boy	.05	.24	.06	.008	.02
	(.005)***	(.009)***	(.006)***	(.001)***	(.008)**
NREGA*Oldest boy	-.02	-.25	-.009	-.006	-.04
	(.006)***	(.01)***	(.008)	(.001)***	(.01)***
# of children in household	-.009	-.01	-.02	-.005	-.02
	(.001)***	(.002)***	(.002)***	(.0003)***	(.002)***
Observations	2,358,118	2,358,118	2,369,967	2,414,033	2,251,174
Mean DV	1.83	2.28	2.73	0.94	4.72
Dist FEs	YES	YES	YES	YES	YES
Year FEs	YES	YES	YES	YES	YES

Notes: This table shows estimates of the effect of NREGA on math and reading test scores and schooling outcomes from the ASER data. All columns contain data from 2005-2009. Math score ranges from 0-3 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide(=3), Math Score2 ranges from 0-5 (Nothing (=0), Recognize Numbers(=1), Can Subtract(=2), Can Divide (=3), Can do first word problem (=4), Can do second word problem 2 (=5), Read score ranges from 0-4 (Nothing(=0), Read Letters(=1), Read Words(=2), Read Paragraph(=3) and Read Story(=4). All Panels include year, district, child age and sex fixed effects. Standard errors clustered at the district-year are reported in parentheses. ***indicates significance at 1% level, ** at 5% level, * at 10% level.