

Essays on Human Capital Formation in Developing Countries



Abhijeet Singh
Nuffield College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Economics
Trinity Term 2014

Abstract

This thesis consists of a short introduction and three self-contained analytical chapters.

Chapter 1 focuses on the question of learning gaps and divergence in achievement across countries. I use unique child-level panel data from Ethiopia, India, Peru and Vietnam to ask at what ages do gaps between different populations emerge, how they increase or decline over time, and what the proximate determinants of this divergence are. I document that learning gaps between the four countries are already evident at the age of 5 years and grow throughout the age trajectory of children, preserving country ranks from 5 to 15 years of age. At primary school age, the divergence between Vietnam and the other countries is largely accounted for by substantially greater learning gains per year of schooling.

Chapter 2 focuses on learning differences between private and government school students in India. I present the first value-added models of learning production in private and government schools in this context, using panel data from Andhra Pradesh. I examine the heterogeneity in private school value-added across different subjects, urban and rural areas, medium of instruction, and across age groups. Further, I also estimate private school effects on children's self-efficacy and agency. I find modest or insignificant causal effects of attending private schools in most test domains other than English and on children's academic self-concept and agency. Results on comparable test domains and age groups correspond closely with, and further extend, estimates from a parallel experimental evaluation.

Chapter 3 uses panel data from the state of Andhra Pradesh in India to estimate the impact of the introduction of a national midday meal program on anthropometric z-scores of primary school students, and investigates whether the program ameliorated the deterioration of health in young children caused by a severe drought. Correcting for self-selection into the program using a non-linearity in how age affects the probability of enrollment, we find that the program acted as a safety net for children, providing large and significant health gains for children whose families suffered from drought.

Notes for the reader

This thesis consists of a brief introduction and three substantive chapters, written to be read independently of each other. The numbering of sections, figures, tables and equations is specific to each chapter. Each chapter is followed by chapter-specific appendices and list of references. Notation may vary across chapters.

Chapters 1 and 2 are solely my own work, Chapter 3 was written collaboratively with Profs. Stefan Dercon and Albert Park.

Chapters 2 has been published as

‘Singh, Abhijeet. Private school effects in urban and rural India: Panel estimates at primary and secondary school ages, *Journal of Development Economics*, vol. 113, pp. 16-32, March 2015.’

Chapter 3 has been published as

‘Singh, Abhijeet, Albert Park and Stefan Dercon 2014. School Meals as a Safety Net: An Evaluation of the Midday Meals Scheme in India, *Economic Development and Cultural Change*, vol. 62(2), pp. 276-305, Jan 2014.’

Word count: This thesis has approximately 56, 918 words excluding the references. This is calculated using the number of words on page 3 (298) multiplied by the number of pages (191).

Acknowledgments

This thesis is largely about investments in childhood. Hence my first acknowledgments are those to my family, especially my parents and my grandfather. Their affection and inputs since infancy, perhaps earlier, surely affect me even today. This thesis would not be possible without them.

I am very grateful to my supervisors, Profs. Stefan Dercon and Albert Park whose encouragement has been of vital importance from the very beginning, including in the decision to return to Oxford for the DPhil. They continued to guide me even after they were both called away from Oxford with competing professional duties. Their approach to research shapes my own (although naturally not the errors, which are solely mine). Most of all, I am grateful that I could always count on them for supporting me at critical junctures throughout my DPhil and rely on them for invaluable advice.

All empirical work in this thesis draws upon data collected by the Young Lives Project at QEH. However, the contribution of Young Lives to this thesis is much greater. My part-time employment there, first as an RA and then as a Research Officer, provided the financial means to undertake the DPhil in the absence of other funding. More than that, Young Lives gave me a collegiate and supportive environment to work in, flexibility which

allowed me to alternately manage doctoral deadlines and pressing work deadlines, and the space to develop my own research interests. Without the support of colleagues at Young Lives, especially Jo Boyden and Graham Bray, I could not have begun the DPhil at Oxford and, having begun, might not have finished it in time. Working with colleagues with a wide variety of research interests and experience has been very rewarding and stimulating; in particular, working with Sofya Krutikova, Caine Rolleston and Zoe James, even on tasks tangentially related to my doctoral work, was very helpful to me at different stages.

I have also benefited greatly from the kindness of many researchers who spared time to comment on my work at seminars and otherwise, read drafts, and provide feedback. In particular, the very constructive feedback and encouragement I received from Prof. John Muellbauer, my College Advisor at Nuffield, and from Prof. Steve Bond and Dr. Francis Teal, who examined me for the Transfer of Status assessment, has shaped many parts of this thesis. Harold Alderman, Jere Behrman, Jishnu Das, James Fenske, Paul Glewwe, Karthik Muralidharan, Lant Pritchett, Simon Quinn and Adrian Wood provided detailed comments on the draft of one or more papers in this thesis, for which I am very grateful. CSAE, QEH and Nuffield provided colleagues and friends who helped improve my research but also provided a close-knit support group which has been an amazing feature of doing the DPhil at Oxford.

Finally, I have deep sense of gratitude to my friends of many years, especially Lee Crawford, Anthony Harris, Aparna John, Vinayak Uppal and Nick Travis. Their support to me personally, their company, and their patient understanding and affection have meant a great deal.

Contents

Introduction	1
1 Evolution of international learning gaps	9
1 Introduction	10
2 Data and context	17
2.1 Data	17
2.2 What are absolute learning levels in these samples? .	21
3 When do gaps emerge and how do they evolve? . .	25
4 What causes divergence? Results from VAMs	31
4.1 Child-specific endowments	31
4.2 Differential productivity of home inputs	37
4.3 Differential schooling exposure and school productivity	42
4.4 Comparison with IV estimates	47
4.5 Robustness checks	52
5 Conclusions	55
Appendices to Chapter 1	66
2 Impact of private schools in India	76
1 Introduction	77
2 Private and Government schools in India	82
3 Data	85
3.1 Sampling	85
3.2 Data collected through household visits	87
3.3 Data collected from schools	90
4 Empirical Framework	91

4.1	Education Production Functions	91
4.2	Are value-added models reliable for the estimation of policy effects?	94
4.3	Estimated specifications	97
4.4	Results	102
5	Additional concerns and robustness	112
5.1	Unobserved differences in ability	112
5.2	Parental information, aspirations and dynamic sorting	115
5.3	Dosage of treatment across samples and school migration	119
5.4	Lag structure	123
6	Do private schools have an impact on psychosocial skills? . .	124
7	Conclusion	128
	Appendices to Chapter 2	140
3	School meals as a Safety Net	144
1	Introduction	145
2	The Midday Meals Scheme in India	151
3	Data and background	153
4	Framework and methodology	156
5	Results	170
6	Conclusion	181
	Appendices to Chapter 3	188

List of Tables

CHAPTER 1	19
1 Enrolment and grade progression of children in Young Lives	19
2 Proportion answering TIMSS questions correctly	22
3 Comparing 12-year old children in Young Lives to TIMSS 4th grade sample	24
4 Linked test scores at 5, 8 and 15 years	26
5 Descriptive statistics of control variables	34
6 Do home factors and child-specific endowments explain divergence?	35
7 Country-specific production functions of achievement: 8 years old	38
8 Country-specific production functions of achievement: 15 years old	39
9 Predicted mean achievement levels under various counterfactual scenarios	41
10 Comparing effectiveness of a grade of schooling: 8-years old	44
11 Comparing effectiveness of a grade of schooling: 15-years old	46
12 Discontinuity-based results on grade effectiveness in Peru and Vietnam	50
Appendix Tables for Chapter 1	66
A.1 List of Items which were split in estimation due to DIF	70
A.2 Allowing non-linearity in lagged achievement: 8-years old	72
A.3 Allowing non-linearity in lagged achievement: 15-years old	73
A.4 Production function estimates at 15-years using Bayesian EAP lagged scores	74
A.5 Estimates correcting for measurement error: 8-years old	75

CHAPTER 2	101
1 Descriptive statistics - Younger Cohort, 8 years (2010) . . .	102
2 Descriptive statistics - Older Cohort, 15 years (2010)	103
3 Cross-sectional differences in test scores (Private - Government)	104
4 Coefficient on Private School dummy	105
5 Heterogeneity by medium of instruction - Rural Areas, 9 years	107
6 Performance on specific English tasks	109
7 Assessing bias due to omitted ability	113
8 Robustness to selection on parental assessments and aspirations - Rural areas	118
9 Years spent in private and government schools between 2007/8 and 2009/10 academic years (inclusive)	120
10 Estimating differential effectiveness of a year of schooling in private vs. government schools	122
11 Statements and aggregate indexes of locus of control and self-efficacy	125
12 The impact of private schooling on psychosocial skills - Rural areas, 9 years	127
Appendix Tables for Chapter 2	140
A.1 Private school effect regressions: 8-years, rural areas	141
A.2 Private school effect regressions: 9-years, rural areas	142
A.3 Private school effect regressions: 15 years, rural areas	143
CHAPTER 3	154
1 Descriptive statistics	163
2 Unconditional treatment effect from OLS regressions on binary treatment variable	171
3 Estimated impact of Midday Meals on children's nutrition .	172
4 Heterogeneous impact of Midday meals on drought-affected children's nutrition	174
5 Results using community-level information on timing and occurrence of drought	177
Appendix Tables for Chapter 3	188

A.1	Descriptive statistics for treatment group, comparison group and all non-beneficiaries	188
A.2	First stage results for endogenous variables	189
A.3	Results using site-averaged drought measure	190
A.4	Estimates using the whole sample	191

List of Figures

CHAPTER 1	18
1 Age of children in Young Lives	18
2 Learning distributions at different ages	28
3 Progress in learning across countries	30
4 Discontinuity in grade attained by month of birth	49
A.1 Item Characteristic Curve	67
A.2 Detecting Differential Item Functioning (DIF)	71
CHAPTER 2	85
1 Timing of interviews in Young Lives	86
CHAPTER 3	154
1 Timeline of events	155
2 Proportion of children enrolled by month of birth	166

Introduction

The study of investments in childhood in developing countries, especially in the domains of health and education, has been in sharp focus in the past decade. This emphasis results, at least in part, from three linked developments over this period. First, a large body of literature, which has grown rapidly, has provided rigorous evidence on the importance of early childhood investments on later life outcomes both in developed and developing countries¹. Second, health and education policies have gained increasing prominence in most developing countries, embodied for instance in the Millennium Development Goals. Third, partly as a result of policy focus but also due to rapid economic growth, the socio-economic contexts determining investments in health and education have changed very rapidly in many developing countries.

This thesis consists of three substantive chapters, each written to be self-contained and read independently of the others, but focusing on research questions within this broad theme of human capital formation in childhood in developing countries. Data used in this thesis were collected

¹This relevant literature will be described in the three analytical chapters. Recent authoritative reviews on some selected aspects may be found in Almond and Currie (2011), Currie and Almond (2011), Heckman and Kautz (2013), Heckman and Mosso (2013), Glewwe and Miguel (2007) and Strauss and Thomas (2007).

by the Young Lives study, a unique panel study of two cohorts of children in Ethiopia, India, Peru and Vietnam, relevant features of which will be explained in the analytical chapters.

Chapter 1 begins with the observation that the performance of developing countries on international assessments of learning varies widely; this spread in the level of cognitive skills may be expected to have longer run consequences, for example to the dispersion of wages and economic growth. With this backdrop, and given the now-large body of evidence on the early origins of divergence in skills and income, I ask at what age do gaps between countries on learning achievement emerge, how they then evolve over the age and educational trajectory of children and what the proximate sources of divergence across countries in learning achievement are; specifically, I am interested in assessing the contribution of differential productivity of schooling to this learning divergence across contexts. The Young Lives data, with identical tests administered across countries to children at 5, 8, 12 and 15 years of age in a panel setting, offers a unique opportunity to study these questions.

Results in Chapter 1 offer several insights. *First*, learning levels in the sample are low relative to international norms: at the age of 12 years, I find that about half the children in Ethiopia, and about a quarter of children in Peru and India fail to reach the low achievement benchmark for fourth-grade children (aged about 10 years) in TIMSS. *Second*, there is a clear stochastic dominance of quantitative proficiency evident at the age of 5 years, before most children have started school, with children in Vietnam and Peru at the upper end, Ethiopia at the lower end, and India in between. This ranking is stable at all ages tested i.e. at 5,

8, 12 and 15 years. Furthermore, there is clear evidence of the gaps increasing over the age range: conditional on test scores in 2006/7, in both cohorts the scores in 2009 are higher in Vietnam than in Peru which are in turn greater than in India and Ethiopia. *Third*, this gap is only partially explained by a rich measure of inputs in a production function approach including household wealth, parental education and the child's nutrition, daily time use, sex and birth order. *Fourth*, the estimated productivity of schooling differs importantly across countries: at primary school level, a year of schooling in Vietnam is considerably more productive in terms of quantitative skill acquisition than a year of schooling in Peru or India. Preferred IV estimates, based on discontinuities in grade completion arising between children born in adjacent months due to country-specific enrolment guidelines, indicate a gain of about 0.4 SD per grade completed in Vietnam in comparison to 0.2 SD in Peru.

Chapter 2 continues on the theme of understanding differences in learning acquisition in childhood, but shifts the focus to study within-country divergence, specifically between students in private and government schools in India. This is currently an issue of great policy and research interest in India. The share of private schools in total enrolment has risen substantially across both urban and rural areas in India in the past 15 years, now accounting for about 30% of enrolment in rural areas; students in these schools perform much better on average in test scores ; and frequently it seems that private schools achieve this better performance even with much lower expenditure per pupil than government schools.

In this chapter, I answer three questions related to the impact of private schools in India. First, I estimate value-added models (VAMs) of learning

achievement to evaluate whether, and to what extent, the better test performance of children in private schools is attributable to schools. Second, I examine whether the impact of private schools on test scores is heterogeneous across different tests, different age groups, different languages of instruction and across urban and rural areas. Third, I attempt to assess if private schools have a differential impact on psychosocial ('non-cognitive') skills of children, specifically their self-efficacy and agency.

In rural areas, I find a substantial positive effect (>0.5 SD) of private schools on English, no effect on mathematics and heterogeneous effects on Telugu for 8–10 year old students; at 15 years, there are modest effects (<0.2 SD) on Telugu, mathematics and vocabulary. I find no evidence of a positive effect in urban areas or on psychosocial skills. Results on comparable test domains and age groups correspond closely with, and further extend, estimates from a parallel experimental evaluation. Combined with previous work highlighting that the average cost per child in rural private schools is less than a third of the average cost in the state schools, and that private schools dedicate less instructional time to Telugu and mathematics, they suggest strongly that private schools are much more productive than government schools. However, they also imply that the spread of private schools is unlikely to raise average achievement levels as measured by mathematics skills or functional literacy significantly; with the exception of English, I do not find any large and consistently positive 'private school effect'.

Chapter 3, the last chapter in this thesis, shifts attention from learning to nutrition. It focuses on the nutritional impacts of the Midday Meals Scheme in India and, specifically, attempts to assess if this program could help

ameliorate the negative nutritional impacts of droughts in early childhood.²

The Scheme provides a cooked meal at school to all children enrolled in government or government-aided schools on every school day; covering over 120 million children by 2006, it is the largest school feeding program in the world. Evaluating these effects offers important challenges. All government schools offered the meals to all children enrolled in primary grades starting at the same time, thus making it hard to find a relevant comparison group.

Importantly for analytical purposes, at the time of the second round of data collection (2007), only around half of the 4-6 year old children had started schooling: so, we could compare children who were already enrolled in government primary schools (and thus getting the Midday Meals) with children who would join these schools later but hadn't yet. Since the age of enrolment into schooling is partly a household choice, it is likely to be endogenous to expected gains from the school meals. To correct for self-selection into the program, we utilize a non-linearity in enrollment induced by a change in the calendar year of birth: this strongly affects the probability of enrollment, as it is used as a 'rule of thumb' to determine the appropriate time for enrollment, but should not directly impact nutrition when controlling for age in the regressions. We use an indicator variable for whether the child was born in 2002 as an instrumental variable (IV) for our treatment dummy variable. Our IV is informative in the dataset, even though the treatment and comparison groups are only about two months apart in age on average, because data collection was carried out just as decisions on school enrollment were being made for the younger cohort children (who were between 4½ and 6 years old in the second round); the

²This chapter is co-authored with Profs. Stefan Dercon and Albert Park.

non-linearity was a sufficiently strong predictor of whether children were in the treatment group at the time of the survey.

We find large benefits for children whose households self-report having suffered from drought between the two rounds; results from our preferred specification suggest drought exerts a substantial negative effect on both nutrition indicators but that these negative effects are entirely compensated for by the Midday Meals Scheme. The magnitude of the effects of the MDMS program is very large: for boys aged 65 months (the mean age in our sample), for example, the preferred specifications using self-reported drought suggest that drought creates a height loss of about 0.77 s.d. which roughly equals the distance between the 25th and 50th percentiles, and a weight loss of about 0.44 s.d. which equals two-thirds of the same distance, in the WHO (2007) growth charts; our results suggest that this entire gap is compensated for by the program.

These results are important in light of the scale and salience of the program in India. Even more so, they are important because they highlight the possibility of ‘catch-up growth’ wherein nutritional deprivations in early childhood may sometimes be remedied even outside the critical window of a thousand days from conception to two years of age. While nutritional deprivations in early childhood are often thought to be irreversible, we discuss with recourse to a large biomedical literature that catch-up may be possible and, importantly from a policy perspective, highlight a large Scheme evaluated at scale which promotes such catch-up. These results are also important given the literature on the effect of environmental shocks on child nutrition in a number of developing countries.

Although the three chapters are intended to be of independent interest, written to be read on their own, there are several commonalities. Most prosaically, they draw upon the same data. Substantively, they focus on questions within the same broad research area - human capital formation in childhood, specifically nutrition and education, in developing countries. They are motivated by the same literature and set of policy priorities (mentioned above) and they focus alike on issues of quality and impact of the provision of childhood investments in these settings. There are strong methodological and substantive complementarities between the chapters: for example, Chapters 1 and 2 both use a value-added modeling approach to focus on similar questions about the learning gains in different educational systems; Chapters 1 and 3 draw upon similar sources of variation, arising from enrolment guidelines and month of birth, for identification; and Chapters 2 and 3, although methodologically different, focus on the same context and study two of the most important developments in the educational sector in India in the past decade.

Thus, despite the self-standing nature of the chapters, I hope that the thesis presents a coherent whole.

References

- ALMOND, D., AND J. CURRIE (2011): “Killing me softly: The fetal origins hypothesis,” *The Journal of Economic Perspectives*, 25(3), 153–172.
- CURRIE, J., AND D. ALMOND (2011): “Human capital development before age five,” *Handbook of labor economics*, 4, 1315–1486.

- GLEWWE, P., E. A. HANUSHEK, S. HUMPAGE, AND R. RAVINA (2013): “School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010,” in *Education Policy in Developing Countries*, ed. by P. Glewwe. University of Chicago Press.
- GLEWWE, P., AND M. KREMER (2006): “Schools, teachers, and education outcomes in developing countries,” *Handbook of the Economics of Education*, 2, 945–1017.
- GLEWWE, P., AND E. A. MIGUEL (2007): “The impact of child health and nutrition on education in less developed countries,” *Handbook of Development Economics*, 4, 3561–3606.
- GRANTHAM-MCGREGOR, S., Y. B. CHEUNG, S. CUETO, P. GLEWWE, L. RICHTER, AND B. STRUPP (2007): “Developmental potential in the first 5 years for children in developing countries,” *The Lancet*, 369(9555), 60–70.
- HANUSHEK, E. A., AND L. WOESSMANN (2008): “The role of cognitive skills in economic development,” *Journal of Economic Literature*, 46(3), 607–668.
- HECKMAN, J. J., AND T. KAUTZ (2013): “Fostering and measuring skills: Interventions that improve character and cognition,” Discussion paper, National Bureau of Economic Research.
- HECKMAN, J. J., AND S. MOSSO (2013): “The Economics of Human Development and Social Mobility,” *unpublished*.
- STRAUSS, J., AND D. THOMAS (2007): “Health over the life course,” *Handbook of Development Economics*, 4, 3375–3474.

Chapter 1

Emergence and evolution of international learning gaps: Panel evidence from four developing countries

Abstract

I study divergence in learning across countries using unique child-level panel data from Ethiopia, India, Peru and Vietnam with linked assessment data at 5, 8, 12 and 15 years of age. Gaps in learning are apparent at 5 (prior to school entry), grow over the age range, and thus preserve country ranks at each age. Children in Vietnam consistently score highest on average followed by Peru, India and Ethiopia respectively. Value-added estimates reveal that early divergence between Vietnam and other countries is largely explained by the differential productivity of a year of schooling in primary school. This result is further confirmed with an IV approach using discontinuities in grade completion arising due to country-specific enrolment guidelines for identification.

1 Introduction

The performance of countries on comparable learning assessments differs dramatically, including among developing countries.¹ Recent research suggests that this differential performance in tests directly translates into difference in growth rates and wage inequality (see Hanushek and Kimko, 2000; Hanushek and Woessmann, 2008, 2012a,b; Kaarsen, 2014).

Results from comparative international tests, based on cross-sectional data on children in middle or secondary school, raise two important questions. First, at what age do gaps between countries emerge and how do they then evolve over the children’s educational trajectories? Second, what are the sources of these gaps - can they perhaps be explained by systematic differences in home inputs, school quality or child-specific attributes such as ability or health across countries? These questions are central to identifying the domains in which interventions may be most required and at what ages.²

I address these questions using a unique child-level panel dataset which tracks two cohorts of children between 2002 and 2009 from four developing countries – Ethiopia, India, Peru and Vietnam. The data are particularly suitable for this analysis: The tests span a long period of childhood from 5 to 15 years; the same tests were administered to children in a particular age group across countries; the panel structure of the data allows me to analyze learning dynamics; and detailed household and background information

¹For example, math scores of 15-year-old students in the 2012 PISA assessment differed by 1.4 standard deviations (SD) between Vietnam and Peru, two of the countries covered in this paper. In comparison, the difference between the US and Finland, the highest scoring Scandinavian country, was 0.38 SD(OECD, 2013).

²Such understanding is especially important in light of a now-large literature which documents that the effectiveness of interventions varies over the age of children (see Heckman and Mosso, 2013 for a recent review).

allows for the estimation of rich production functions of achievement. Since the data were collected through home visits of a random sample of children in given birth cohorts, they do not suffer from selection issues arising from non-enrolment or non-attendance that are characteristic of school-based assessments in developing countries, especially in post-primary education. The data are also particularly well suited to this analysis since the four countries display remarkably different levels of achievement.

Using these data, I generate comparable distributions of test scores for quantitative proficiency across countries at the age of 5, 8, 12 and 15 years; I further link the test scores of 12-year old children to the international test distribution in mathematics from the Trends in International Mathematics and Sciences Study (TIMSS) 2003 round which covered Grade 4 students in 29 countries. I compare the achievement distributions for the four country samples in the Young Lives data at each age and assess if there is a clear ranking across the samples, the magnitude of the gaps between countries, and the stability of rankings across different ages.

Further, I use the individual-level panel dimension of the data, combined with household-level information, to estimate value-added models of achievement production. Specifically, I focus on assessing whether test score divergence across countries is explained by differing child-level endowments and home investments, differing time use patterns, and differing exposure to and effectiveness of a grade of schooling in the four countries. For two countries, Peru and Vietnam, I am able to use discontinuities in the number of grades completed arising from enrolment thresholds for the identification of grade productivity effects and compare results to value-added estimates; I find little evidence of systematic bias

in the value-added estimates compared to instrumental variable estimates based on this discontinuity.

This paper makes two key contributions. Foremost, it presents the first analysis of the emergence and evolution of gaps in cognitive achievement across countries, from the age of 5 to 15 years, using internationally comparable child-level panel data. While similar exercises have previously been carried out within the context of individual developed countries while studying socioeconomic or racial gaps in test scores (see e.g. the voluminous literature on the Black-White test score gap in the US), I am not aware of any research that attempts to link analyses comparably across countries or covers an equivalently long age range, even within-country, using data from developing countries. Secondly, it provides the first comparable causal estimates of school productivity across countries which are identified using individual-level panel data combined with quasi-experimental variation.³

Results from the analysis are informative and often striking. Descriptively, learning levels in the sample are low relative to international norms (TIMSS) in all countries except Vietnam, gaps between countries are evident at an early age prior to school entry, and grow in magnitude over the age range studied. A clear pattern of stochastic dominance of quantitative proficiency is evident across ages with children in Vietnam consistently scoring highest followed by Peru, India and Ethiopia respectively. There is, moreover, clear evidence of the gaps increasing in magnitude over the

³A small body of previous work has attempted to estimate the differential quality of schooling across countries, typically relying on either the returns to schooling of immigrants from different countries to the US or on cross-sections of academic achievement from international testing programs (Hanushek and Kimko, 2000; Hendricks, 2002; Schoellman, 2012). These measures relate more to the (quality-adjusted) stock of human capital rather than the productivity of schooling directly which is the parameter of interest in this paper.

age range: conditional on test scores in 2006/7, in both cohorts the scores in 2009 are higher in Vietnam than in Peru which are in turn higher than in India and Ethiopia.

The divergence in learning after 5 years is only partially explained by a rich measure of inputs in a production function approach including household wealth, parental education and the child's nutrition, daily time use, sex, birth order and lagged achievement. In primary school years, I show that the estimated productivity of schooling differs importantly across countries: at primary school level, a year of schooling in Vietnam is considerably more productive in terms of quantitative skill acquisition than a year of schooling in Peru or India. Preferred IV estimates indicate a gain of about 0.4 SD per grade completed in Vietnam in comparison to 0.2 SD in Peru. Equalizing this measure of schooling productivity ('learning-gains-per-grade-completed') in other countries to Vietnamese levels closes entirely the gap between Vietnam and Peru, and 70% of the gap with India, even keeping all other inputs including learning levels at 5 years at their initial levels.

These results have wide-ranging relevance. The finding that the ranking of the four country samples in this paper is evident even at the age of 5 years, before children have begun schooling, supplements a much broader literature across disciplines in providing suggestive support for an increased focus on preschool interventions and early childhood interventions (see e.g. Grantham-McGregor, Cheung, Cueto, Glewwe, Richter, and Strupp, 2007). While a large body of evidence in this area is available from developed countries, with the exception of nutritional or health

interventions, the literature on causal studies of preschool factors on cognitive achievement remains limited in other contexts.⁴

However, equally importantly, I document that the gaps magnify over time with divergence in learning levels at primary school ages, especially between Vietnam and the other countries, largely being explained by the differential productivity of schooling. This indicates that there may be considerable room for corrective policy measures aimed at narrowing learning gaps in primary schools. Foremost, this highlights the need for a shift towards emphasizing learning outcomes in developing countries rather than merely enrolment or grades completed; learning levels in these countries are low and there are important differences in school productivity between different countries, which may offer margins for policy improvement.⁵ The results also emphasize that the ranking of countries by national income (GNI) per capita and by learning outcomes are not necessarily identical: whereas Vietnamese GNI per capita (PPP) is at a similar level as India, and only about a third of Peru, learning outcomes in Vietnam are consistently better than any of the other countries.

⁴For notable exceptions using Latin American data see Berlinski, Galiani, and Manacorda (2008) and Berlinski, Galiani, and Gertler (2009). There is also a broader inter-disciplinary literature (see e.g. Engle, Black, Behrman, Cabral de Mello, Gertler, Kapiriri, Martorell, and Young 2007) but it is mostly associational.

⁵For a detailed discussion of why policy needs to increasingly focus on learning goals rather than enrolment or inputs, see Pritchett (2013). This shift in priorities is increasingly embodied in recent policy discussions (see e.g. Muralidharan, 2013 for India) including discussions on the formulation of international development targets after the expiration of the Millennium Development Goals in 2015 (UN, 2013). This change in focus is especially relevant now given that enrolment is high in most countries and a large body of evidence has emerged that improvements in school inputs, as opposed to pedagogy or teaching reforms, have a very weak relationship to learning improvements (see e.g. Glewwe, Hanushek, Humpage, and Ravina 2013; Kremer, Brannen, and Glennerster 2013; Das, Dercon, Habyarimana, Krishnan, Muralidharan, and Sundararaman 2013; McEwan 2013). That looking at quantity of schooling alone may be misleading was emphasized in an early contribution by Behrman and Birdsall (1983).

Finally, the stark differences in the productivity-per-school-year across countries raise a very important question: ‘Why is learning-productivity-per-year so much greater in some countries than others?’ Most current work within the economics of education in developing countries focuses on the effect of particular interventions (e.g. provision of textbooks) within specific contexts; while this is most useful in allowing for robust identification of policy levers that are available to national governments, it is not adequate for assessing how learning gains in a ‘business-as-usual’ sense differ across contexts – yet there may be important policy lessons to be gained also from asking the latter question.

The analysis and results presented here relate to several strands of the literature within economics. Methodologically, this paper relates closely to the literature on the emergence and evolution of test score gaps between different racial groups or gender (see e.g. Fryer and Levitt, 2004, 2006, 2010, 2013; Todd and Wolpin, 2007), on value-added models of achievement using household-based panel data (see e.g. Todd and Wolpin, 2003, 2007; Fiorini and Keane, 2014) and on mapping test scores on a comparable metric using Item Response Theory (IRT) models.⁶

Substantively, in focusing on patterns and determinants of skill acquisition *at the child level*, the analysis relates directly to a large and rapidly growing microeconomic literature on the economics of education in developing countries that documents low levels of learning (see Glewwe and Kremer, 2006 for an authoritative review), studies trajectories

⁶These models are commonly used in comparative educational assessments (see e.g. Van der Linden and Hambleton, 1997; Mullis, Martin, Gonzalez, and Chrostowski, 2004; OECD, 2013) but are rare within development economics (for notable exceptions, see Das and Zajonc, 2010; Andrabi, Das, Khwaja, and Zajonc, 2011).

in skill acquisition (see e.g. Schady et. al. 2014), and experiments with different interventions to improve these low levels of learning (see Kremer, Brannen, and Glennerster 2013 and McEwan 2013 for meta-analyses).

However, in focusing primarily on the gaps *across countries* in the stock of human capital (learning levels) and in the productivity of a grade of schooling, this paper also has direct relevance for two strands of the applied macroeconomics literature — studies seeking to explain the effect of differential human capital and schooling quality to growth (Hanushek and Kimko, 2000; Hanushek and Woessmann, 2008; Schoellman, 2012; Kaarsen, 2014); and studies seeking to study differences in productivity across countries in different sectors.⁷ Results on differential grade productivity may indirectly relate to differences in the quality of management across different countries (Bloom and Van Reenen, 2007, 2010; ?).

The rest of this paper is structured as follows: Section 2 describes the data used in this paper and puts the samples in context by comparing achievement with the international test distribution in TIMSS 2003; Section 3 investigates stochastic dominance of test outcomes across age groups and assesses whether learning gaps seem to narrow or widen over time between countries; Section 4 estimates value-added models of achievement, presenting assessments of school effectiveness and of inter-group differences in achievement; it also presents some sensitivity

⁷See, for example, Hsieh and Klenow (2009) who study factor productivity and misallocation across Chinese and Indian firms or Gollin, Lagakos, and Waugh (2014) who investigate the agricultural productivity gap in developing countries using household survey data.

analyses for robustness of results; Section 5 discusses the findings and concludes.

2 Data and context

2.1 Data

This paper uses data collected by the Young Lives study in Ethiopia, India (Andhra Pradesh state), Peru and Vietnam which has tracked two cohorts of children over multiple rounds since 2002. The older cohort (‘OC’ hereafter, born in 1994/95) was aged about 8 years and the younger cohort (‘YC’ hereafter, born in 2001/2) was aged between 6-18 months at the time of the first wave of the survey in 2002. In each country, 2000 children of the younger cohort and 1000 children in the older cohort were surveyed.⁸ Two subsequent waves of household-based data collection were carried out in 2006 and 2009. The data are clustered and cover 20 sites in each country across rural and urban areas.⁹ In cases where children have moved from the original communities they were surveyed in since 2002, the study tracked them to their new location. As a result attrition in the data is very low with more than 90% of the original sample still in the survey in 2009 in each country. In this paper, I use test data on quantitative skills for both

⁸The only exception is Peru where only 716 children in the older cohort were surveyed due to resource constraints.

⁹Sites correspond to sub-districts in Ethiopia (kebeles), India (mandals) and Vietnam (communes) and to districts in Peru. Sites were chosen purposively to reflect the diverse socio-economic conditions within the study countries and therefore are not statistically representative for the country: comparisons with representative datasets like the DHS samples do show however that in each of the countries, the data contain a similar range of variation as nationally representative datasets (Outes-Leon and Sanchez, 2008; Kumra, 2008; Escobal and Flores, 2008; Nguyen, 2008).

cohorts from the 2006 and 2009 rounds.¹⁰ Figure 1 presents the age of children in the two cohorts at each survey wave¹¹.

Figure 1: Age of children in Young Lives

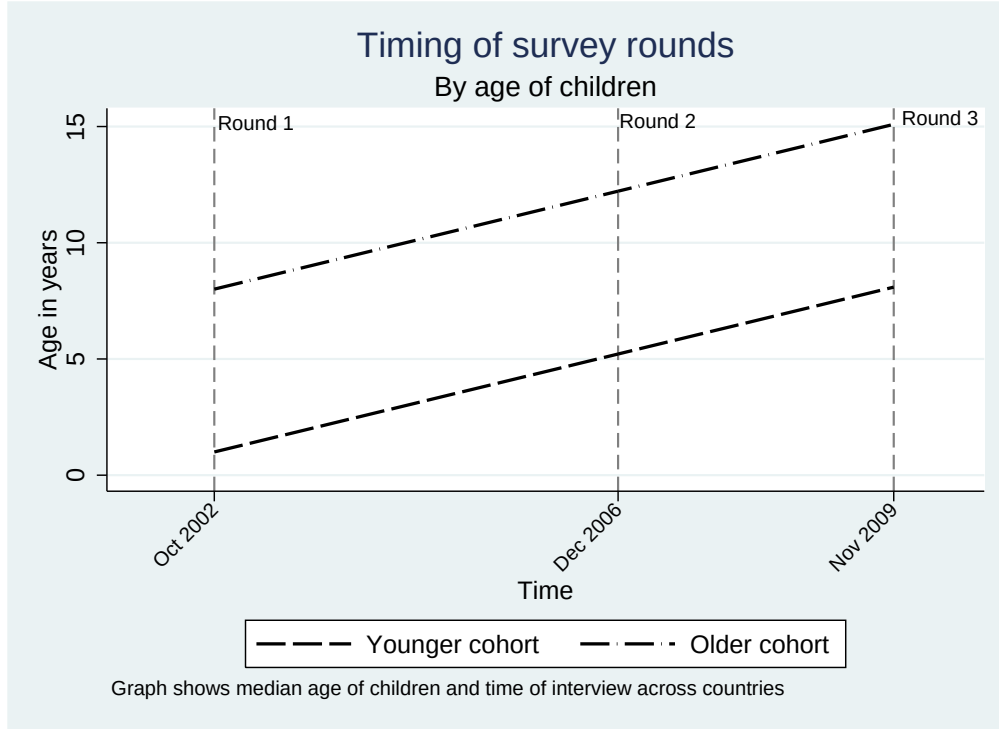


Table 1 presents descriptive information about the educational trajectories and progression of children in the different age samples in Young Lives at different ages which will be of central importance in interpreting all results in this paper. Three patterns from this Table are worth highlighting: ever-enrolment is high across all countries with nearly all children having been enrolled at some point in primary school; the age of entry varies importantly between countries, being the lowest in India (with about 44% of children already in school by 5 years of age) and highest in Ethiopia (with

¹⁰The 2002 round had limited achievement data. On quantitative proficiency, only a single test item ('2 x 4 = ?') was administered.

¹¹Fieldwork typically took between 4-6 months in each country in each round. The timing of the survey rounds shown in Figure is thus only indicative.

an average above 7 years); and the rates of drop-out by the age of 15 also importantly vary with the highest dropouts being in India and Vietnam where just under a quarter of the 15-year old sample is no longer enrolled in school.

Table 1: Enrolment and grade progression of children in Young Lives

Cohort	Variable	Statistics	Ethiopia	India	Peru	Vietnam
Younger cohort (YC)	Age of starting school	Mean	6.9	5.78	6.04	6.05
		SD	0.85	0.82	0.42	0.27
Older cohort (OC)	Age of starting school	Mean	7.19	5.04	5.88	6.07
		SD	1.52	0.71	0.57	0.48
YC 2006 (5-years)	Enrolment	Mean	0.04	0.45	0.01	0.01
		SD	0.19	0.5	0.1	0.08
YC 2009 (8-years)	Enrolment	Mean	0.77	0.99	0.98	0.98
		SD	0.42	0.1	0.13	0.13
OC 2006 (12-years)	Enrolment	Mean	0.95	0.89	0.99	0.97
		SD	0.22	0.32	0.1	0.18
OC 2009 (15-years)	Enrolment	Mean	0.89	0.77	0.92	0.77
		SD	0.31	0.42	0.27	0.42
YC 2009 (8-years)	Highest grade completed	Mean	0.64	1.63	1.31	1.71
		SD	0.77	1	0.58	0.57
OC 2006 (12-years)	Highest grade completed	Mean	3.17	5.61	4.91	5.57
		SD	1.68	1.25	1.11	0.94
OC 2009 (15-years)	Highest grade completed	Mean	5.55	8.15	7.72	8.29
		SD	2.05	1.73	1.31	1.25

Younger cohort ('YC') were born in 2001/2002. Older cohort ('OC') were born in 1994/95

Age of starting school is summarized in both cohorts over those individuals who have enrolled in school at some point before the survey round in 2009.

In each round, a range of background information and child-specific data, including cognitive and nutritional outcomes, was collected.¹² In 2006, quantitative skills were tested for the younger cohort (then aged about 5 years) using the 15-item Cognitive Developmental Assessment tool developed by the International Evaluation Association Preprimary Project; a 10-item Mathematics test was administered to the older cohort, then aged about 12 years, which included six math problems from the publicly released items from the fourth grade tests of the Trends in International

¹²Summary statistics on these variables will be presented in Section 4 at the point they are being introduced in value-added specifications of achievement production.

Mathematics and Science Study (TIMSS) 2003 round. In 2009, a 29-item mathematics test was administered to the younger cohort and a 30-item test to the older cohort.¹³ The survey instruments, including tests, were harmonized across countries in each round.

The analysis in this paper requires generating comparable test scores for each cohort/round sample and further, for the 12-year old sample, scores that are directly comparable to those from the TIMSS 4th grade sample. This is achieved using Item Response Theory (IRT) models which are estimated as in Das and Zajonc (2010) who linked responses to mathematics questions administered in two states of India to the TIMSS 8th Grade test. These models provide several advantages for the purpose of this analysis: first, by explicitly mapping the relationship between the probability of answering a particular test item correctly and an individual's ability, they provide a less arbitrary aggregate measure of proficiency than a percentage correct score which assigns equal weight to all questions, regardless of difficulty; second, they allow for linking across samples even with only partial overlap of test items; and third, they provide for a more robust framework for diagnosing comparability in item performance across contexts, which may be violated due to, for example, translation issues or cultural specificity of items.¹⁴

Item Response models only identify ability (θ) up to a linear transformation i.e. any transformation of the form $a + b\theta$ is an equally valid test score.

¹³See Cueto, Leon, Guerrero, and Muñoz (2009) and Cueto and Leon (2013) for details of the psychometric testing in Young Lives across different rounds.

¹⁴The use of IRT models is standard in educational assessments to generate test scores that are comparable over time or across different populations; it is used, among other applications, in the generation of test scores for the GRE, SAT, TIMSS, PISA and NAEP in the US.

This implies that in the absence of common items, which can be used to ‘anchor’ estimates in two samples, we cannot compare the absolute levels of achievement across two samples. Since tests were not harmonized across rounds or across cohorts in Young Lives, I cannot link test scores on a comparable scale over time or across cohorts. Appendix A provides a brief explanation of IRT and details the procedures for the generation of test scores, as well as analyses to check for and accommodate instances of the differential performance of items across the four countries.

2.2 What are absolute learning levels in these samples?

In this section, I compare the performance of the 12-year old sample in the Young Lives data to the performance of fourth grade students covered by the TIMSS assessment in 2003 who were administered a subset of the same questions. This is useful for putting these samples in context but also is valuable information in its own right: none of these countries have been covered by the TIMSS 4th Grade assessment previously although some comparisons at 15 years of age exist.¹⁵

¹⁵The PISA assessment at the age of 15 years has previously covered Peru (in 2000, 2009 and 2012), Vietnam (2012) and two states in India (2009), although not Andhra Pradesh the state covered by the Young Lives study. Ethiopia has never been covered by TIMSS or PISA.

Table 2: Proportion answering TIMSS questions correctly

Item details	Q1	Q2	Q3	Q4	Q5	Q6
TIMSS Item Number	M011007	M011024	M011011	M011004	M031310	M031162
Cognitive Domain	Using Concepts	Knowing Facts and Procedures	Solving Routine Problems	Using Concepts	Solving Routine Problems	Using concepts
<i>TIMSS</i>						
Canada - Quebec	0.925	0.926	0.847	0.894	0.690	0.636
England	0.929	0.959	0.875	0.858	0.787	0.818
Hong Kong	0.977	0.983	0.916	0.848	0.950	0.746
Italy	0.924	0.968	0.831	0.850	0.731	0.785
Japan	0.968	0.988	0.928	0.895	0.893	0.900
Morocco	0.699	0.775	0.457	0.475	0.441	0.491
Netherlands	0.976	0.949	0.949	0.896	0.893	0.864
Philippines	0.633	0.859	0.573	0.600	0.282	0.416
Singapore	0.974	0.970	0.936	0.895	0.941	0.878
Tunisia	0.813	0.739	0.609	0.495	0.622	0.376
USA	0.929	0.936	0.884	0.887	0.674	0.666
<i>Young Lives</i>						
Ethiopia	0.606	0.716	0.512	0.504	0.404	0.574
India (A.P.)	0.737	0.822	0.603	0.682	0.385	0.711
Peru	0.701	0.912	0.679	0.769	0.508	0.650
Vietnam	0.843	0.938	0.759	0.687	0.751	0.845

Questions focused on the number content domain and were taken from the released items for the 2003 TIMSS assessment.

The TIMSS sample is from Grade 4, aged 10 years on average, at the time of testing in 2003. Cells contain unweighted proportions.

The Young Lives sample is for the older cohort, aged about 12 years at time of testing in 2006.

Table 2 presents the proportion of test-takers in the Young Lives sample, and in the TIMSS 4th Grade sample from select countries which had participated in the 2003 round, who answered the six common questions correctly. As is clear from the Table, a smaller proportion of 12-year old children in the Young Lives sample answer the common test questions correctly than 10-year old children in most OECD economies. The only exception is Vietnam where the proportion correct across questions seems to be comparable to many OECD country samples and is significantly higher than in the other Young Lives countries for most questions.

Comparing the full learning distribution across samples, whether across the Young Lives countries or to the TIMSS sample, requires the aggregation of responses to different test items into a single test score. In order to allow such comparisons, I use the six common items from TIMSS as anchor

items and generate comparable IRT test scores.¹⁶ Table 3 presents the median of the distribution of mathematics achievement in the 12-year old sample and the proportion of children who fall below the low and intermediate benchmarks developed by TIMSS and linked to their test scale for mathematics.¹⁷

The median child in Ethiopia, India and Peru is below the median in the TIMSS 4th Grade sample and with significant gaps in proficiency compared to children in OECD economies. More meaningfully, over a quarter of the sample in India and Peru, and about half in Ethiopia, fail to reach even the Low International Benchmark, defined by TIMSS as follows: *“Students have some basic mathematical knowledge. Students demonstrate an understanding of whole numbers and can do simple computations with them.”* (Mullis, Martin, Gonzalez, and Chrostowski, 2004). In Peru and India, the median child is behind the intermediate benchmark for fourth-graders even though he is in 6th and 7th grades respectively in the two countries; despite 2-3 additional years of schooling, this is notably worse than for any of the selected countries from the TIMSS sample displayed in Table 3.

¹⁶Item characteristics, i.e. the item specific parameter values, were taken from the TIMSS 2003 report for the common items and treated as known with the item characteristics of the other questions and the ability of the sample individuals treated as unknown parameters to be estimated. TIMSS reports test scores by rescaling the proficiency estimates to have a mean of 500 and a standard deviation of 100 for their international sample; I follow the same procedure in order to keep test scores exactly comparable to the TIMSS sample.

¹⁷TIMSS defines four international benchmarks – Low (400), Intermediate (475), High (550) and Advanced (625) – each with descriptions of the level of skills. See Mullis, Martin, Gonzalez, and Chrostowski (2004) for details.

Table 3: Comparing 12-year old children in Young Lives to TIMSS 4th grade sample

Country	Median grade	Median score	% below low benchmark (400)	% below intermediate benchmark (475)
<i>Young Lives sample (12 years old, 2006)</i>				
Ethiopia	4	394	52	82
India	7	460	28	54
Peru	6	454	25	59
Vietnam	7	525	13	38
<i>TIMSS 4th grade sample (selected countries, 2003)</i>				
Australia	4	504	12	36
Canada - Quebec	4	509	6	30
Chinese Taipei	4	567	1	8
England	4	536	7	25
Hong Kong	4	578	1	6
Italy	4	507	11	34
Japan	4	568	2	11
Netherlands	4	542	1	11
Russian Federation	4	533	5	24
Singapore	4	601	3	9
United States	4	522	7	28

Test scores are IRT scores linked to TIMSS 2003 assessment using item parameters of publicly released items for anchoring and normalized as in TIMSS.

TIMSS normalizes scores to have a mean of 500 and standard deviation of 100 in the international pooled sample. TIMSS sample performance taken from Mullis et. al. (2004).

In interpreting these IRT estimates, it is useful to point out that the assessment in this age group is affected importantly by ceiling and floor effects, thus not capturing the full spectrum of ability.¹⁸ Since the mean of the distribution may be affected by these ceiling and floor effects, I have

¹⁸About 33% of children in Vietnam, 18% in India, 8% in Peru and 4% in Ethiopia answered all ten questions correctly while about 8% of children in Ethiopia, 6% in India, 3% in Vietnam and 1% in Peru answered none correctly. IRT maximum likelihood estimates are undefined when respondents answer fewer questions correctly than would be expected by guessing or answer all questions correctly. The common solution adopted in practice to address this issue is to bound the upper and lower limit of the ability distribution. For example, TIMSS and Das and Zajonc (2010) bound the ability distribution between [5,995] and until their recent revision of scales, GRE scores were bounded between [200,800]. For this sample, scores are bounded between [0,1000].

restricted myself to only comparing the performance of the median child for this age sample.¹⁹

3 When do gaps emerge and how do they evolve?

Table 3 and the foregoing discussion highlight two patterns quite starkly: that there exist vast differences in the mathematical knowledge acquired by children of the same age across the four countries in the Young Lives samples; and that the absolute levels of learning, with the possible exception of Vietnam, are very low in comparison to OECD countries, even though in Peru and India the children have had two extra years of schooling compared to the OECD sample. In this section, I analyse whether a similar pattern is also discernible at other ages and whether there are not only systematic levels differences between children in different countries but also differences in how much they learn over time.

Table 4 presents descriptive statistics of the quantitative ability scores at each of the other three ages at which testing was done i.e. at the age of 5 years, 8 years, and 15 years. The test scores are comparable across countries at each age and are normalized internally to have a mean of 500 and a standard deviation of 100 at each age; test scores are not linked across cohorts or over time and therefore cannot be directly compared across age groups.

¹⁹This bounding does not affect the qualitative results presented in this section which are also borne out by plausible value estimates not subject to the same concerns. For the other age-samples presented in this paper, this bounding exercise affects only a small proportion of the sample as the tests provide a smooth measure of achievement with no ceiling or floor effects. The censoring of achievement for the 12-year old sample will also present challenges for the estimation of value-added models later since the inclusion of censored regressors can induce bias; I will test for this concern explicitly in Section 4.

Table 4: Linked test scores at 5, 8 and 15 years

Age group	Statistics	Countries				Total
		Ethiopia	India	Peru	Vietnam	
5 years	Mean	454	498.3	520.4	524.7	499.8
	p25	402.4	442.1	462.9	472.6	442.9
	p50	456.1	491	511	522.7	495.3
	p75	503.8	540.7	569.1	575.6	550.8
	SD	102.1	94.8	97.6	89.1	99.9
	N	1846	1904	1893	1935	7578
8 years	Mean	419.1	495.9	518.2	563.6	500
	p25	377.7	458.2	491.4	535.9	456.5
	p50	426	502	530	570	518.7
	p75	488.8	542.3	555.8	600.5	559.4
	SD	100.7	84.6	68.3	85.3	100
	N	1885	1930	1943	1964	7722
15 years	Mean	442.7	482.1	527.7	556	500
	p25	433.1	467	512.9	528	478.3
	p50	480.1	508.4	535.9	561.6	522.6
	p75	512	543.1	554.6	590.7	555.7
	SD	106.5	99.6	54.9	79.7	100
	N	974	977	678	972	3601

Scores are IRT test scores generated within an age sample, pooling data from all countries, and normalized to have a mean of 500 and an SD of 100 in the pooled sample. Scores are comparable across countries but not across age groups.

The most striking pattern in Table 4 is that the ranking of countries seen at the age of 12 years is already evident at the age of 5 years, which pre-dates schooling for most of the sample.²⁰ The second notable pattern is that the ranking of countries in the Young Lives sample is unchanged across age groups. Although occasionally differences between countries are not statistically significant – at the age of 5 between Vietnam and Peru, and

²⁰Given the literature from OECD countries about early emergence of cognitive gaps, this is perhaps unsurprising. However, this pattern is notable because analysis on the cognitive impact of environmental influences which precede school enrolment (with the exception of nutrition or health shocks, (e.g. Glewwe, Jacoby, and King, 2001 and Maccini and Yang, 2009) remains very limited in developing countries.

at the age of 12 between Peru and India – the general ranking of the four country samples is remarkably stable. This ranking is also the same as the ranking implied by the PISA test scores where Vietnamese children score significantly higher than children in Peru and Peruvian children score significantly higher than Indian children.

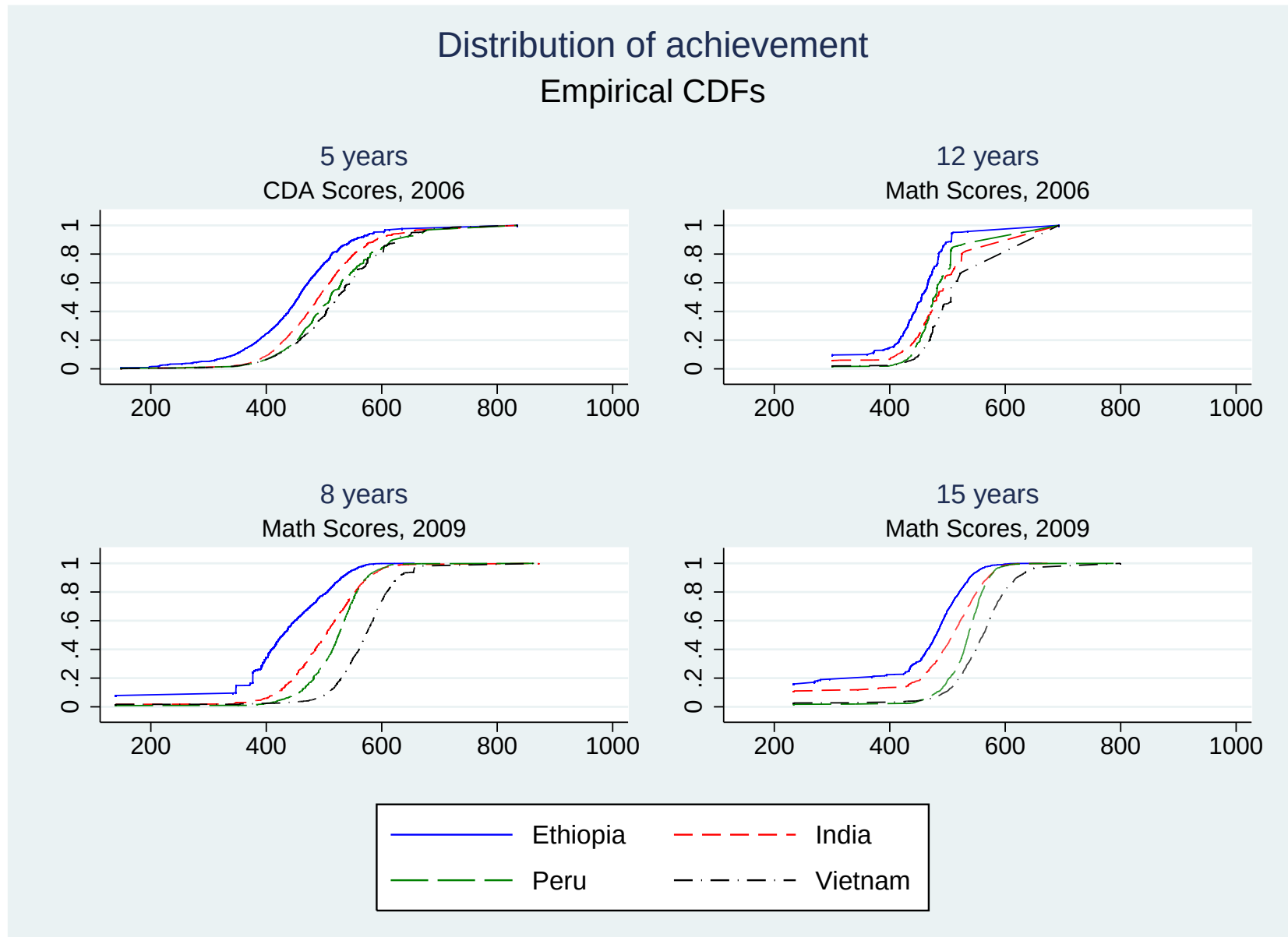
Conclusions formed on the basis of the mean comparisons also hold true across the entire distribution. The CDFs of the estimated test scores are plotted for each of the four age groups in Figure 2: there is a clear pattern of stochastic dominance with Vietnamese children performing better at every age compared to other samples and with Ethiopian children performing distinctly worse.²¹ With the anomalous exception of the 12-year sample, the Peruvian CDF always lies to the right of the Indian CDF. In general, there seems to be a clear and stable ranking with Vietnam > Peru > India > Ethiopia in these samples.²²

Since test scores are not linked across age groups, Figure 2 is not enough by itself to comment on whether differences between country samples are exacerbated as children grow older. Further, even if gaps have grown, Figure 2 does not answer whether any further divergence is only caused by amplification of initial gaps (through the self-productivity of skills) or through other channels in achievement production. In Figures 3a and 3b,

²¹In generating Figure 3, and in all subsequent analysis in this paper, I have re-normalized the 12-year old achievement to have a mean of 500 and SD of 100 in the pooled Young Lives sample (rather than linked to the TIMSS normalization as in Section 2) in order to keep normalization procedures identical for each age group. Ceiling and floor effects in the 12-year old sample can be noted in the empirical CDFs.

²²This is important because mean comparisons are not adequate to make judgments about the entire distribution of learning across countries. As Bond and Lang (2013) point out, citing Spencer (1983), the ordinality and arbitrary normalization of test scores implies that the only way of reliably ranking samples is to look at the cumulative distribution functions (CDFs) of achievement.

Figure 2: Learning distributions at different ages



CDFs show distribution of test scores estimated with Item Response models pooling all country samples in each age group. Scores are internally normalized to have a mean of 500 and standard deviation of 100 in each age sample.

I present non-parametric plots of achievement in 2006 and achievement in 2009 for the four country samples in both cohorts. The essential idea behind these graphs is simple: conditional on test scores in 2006, do we see children in the four countries achieve similar results in 2009 (in which case gaps at later ages only reflect past divergence), or do we see children in some countries perform better than children in other countries who had scored similarly in 2006 (in which case there is additional divergence)?

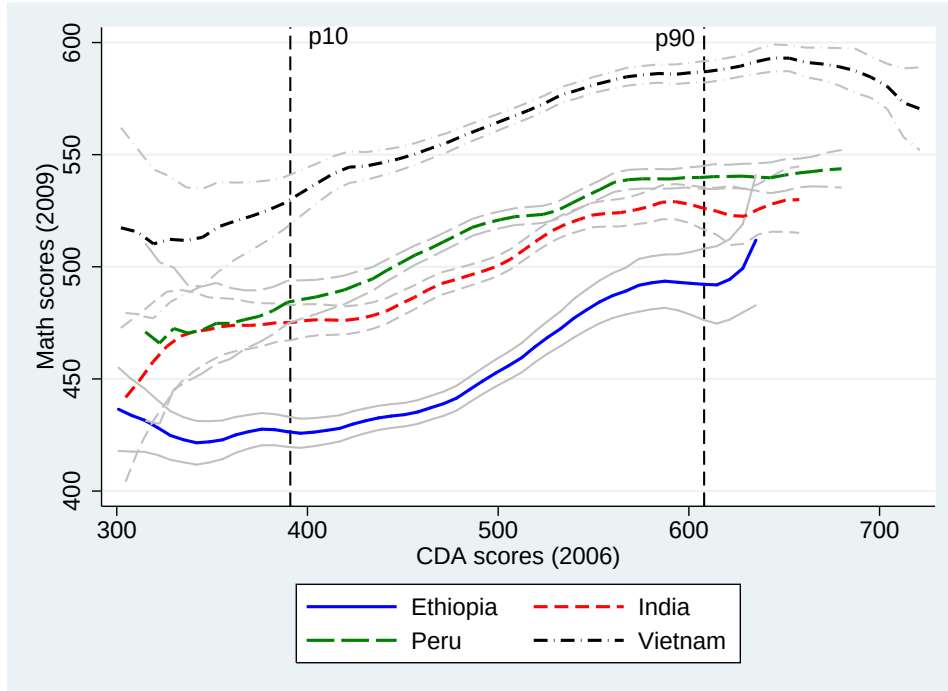
Between 5 and 8 years of age (Figure 3a), there seems to be considerable divergence between countries which is similar to the ranking of the country samples on the levels of achievement at age 5: children in Vietnam learn more than children in Peru, who in turn learn more than children in India and Ethiopia respectively, even conditional on having achieved the same score at age 5. Between the age of 12 and 15 (Figure 3b), however, a somewhat different picture emerges: children in Vietnam and Peru display near-identical trajectories of achievement which are higher than children in Ethiopia and India which are almost indistinguishable from each other. ²³

The difference between the two sets of countries seems to be a difference in the intercepts and not the slopes which suggests that divergence between ages is not due to the initial learning levels in the four samples.

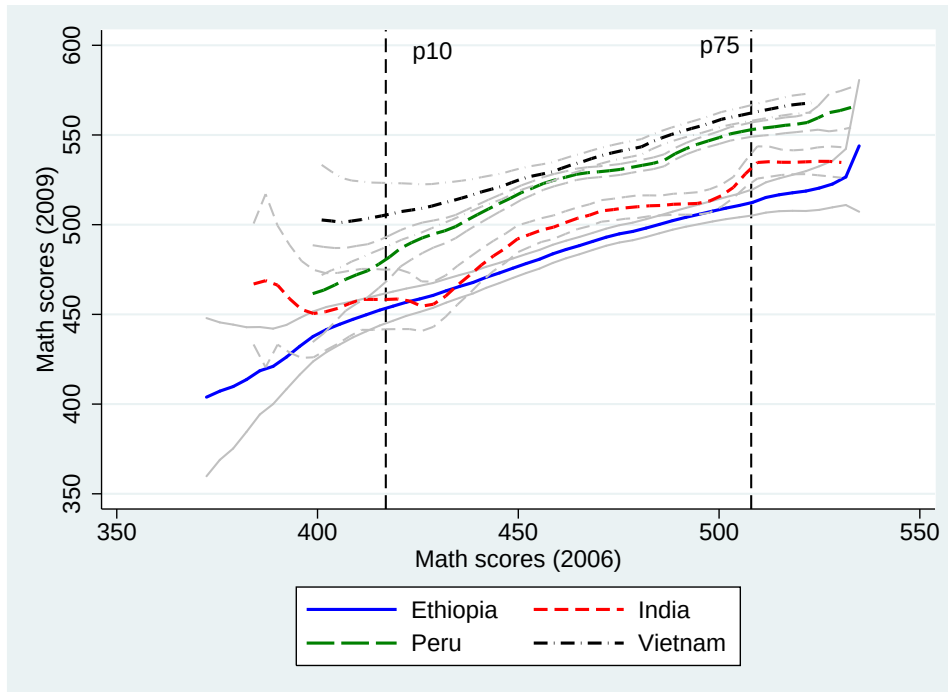
²³Similar patterns of divergence in these data are also documented in Rolleston (2014) and Rolleston, James, and Aurino (2013). The most important difference between their analysis and this paper is in the use of IRT models here which offer a better conceptual basis for cross-cultural comparison, allow different test items to contribute differently to the aggregate test score and provide a more continuous measure of ability in comparison to percentage correct scores as used by both Rolleston (2014) and Rolleston, James, and Aurino (2013). Despite differences in method, however, the basic descriptive findings are similar. Rolleston (2014) and Rolleston, James, and Aurino (2013) do not attempt comparisons of individual performance in the Young Lives sample with the performance of the international test distribution in TIMSS or an analysis of the sources of divergence across countries.

Figure 3: Progress in learning across countries

(a) 5-8 years



(b) 12-15 years



Note: Lines are local polynomial smoothed lines shown with 95% confidence intervals and reference lines for relevant quantiles.

The sample is restricted to observations not suffering from ceiling or floor effects in either round.

4 Sources of divergence: Results from value-added models

Analysis presented in the previous section documents the pattern of divergence across different country samples but is only partly informative about the sources of this divergence. Documenting patterns of learning even conditional on past achievement is insufficient by itself to say, for example, whether the divergence is primarily a factor of school inputs or a result of constant application of superior home inputs at every life stage in some contexts than others; from a policy perspective, however, identifying sources of divergence is of considerable interest. In this section, I estimate value-added models of achievement production to address this issue.

4.1 Do child-specific endowments explain divergence?

As a benchmark case, I first explore sources of achievement across the four countries at each age group as follows:

$$Y_{ic,a} = \alpha + \beta_1 \cdot \theta_c \tag{1}$$

$$+ \beta_2 \cdot Y_{ic,a-1} \tag{2}$$

$$+ \beta_3 \cdot X_{ic} \tag{3}$$

$$+ \beta_4 \cdot TU_{ic,a} + \epsilon_{ica} \tag{4}$$

where $Y_{ic,a}$ is the test score of child i in country c at age a ; θ_c is a vector of country dummy variables (with Ethiopia as the omitted category); X_{ic} is a vector of child-specific characteristics which includes caregiver's education (in completed years), child's age in months at time of testing, child's height-for-age z-scores at time of testing (based on WHO 2005 standards), a wealth index based on durables owned by household and access to services and dummy variables for being male and being the eldest child; TU is a vector controlling for time use across different tasks on a typical day (with sleeping being the omitted category). The estimation is carried out separately for the two cohorts (i.e. at ages 8 and 15) by pooling all country samples within cohort.

Inclusion of controls is sequential as detailed in Equations 1-4 and naturally changes the interpretation of coefficients. Specification (1) displays mean difference between countries at the ages of 8 and 15 years; Specification (2) is the linear regression analogue of Figure 3 and shows the divergence between the countries, conditional on lagged individual test scores; Specifications (3) and (4) further explore if the divergence is explained by the levels of covariates in X or in current time use respectively. Time use is entered in the final step since it potentially conflates (through the categories of time spent at school or studying after school) school inputs with home-based inputs.²⁴

²⁴The inclusion of time use categories should thus be considered here in the spirit of a bounding exercise, exploring the upper bounds of how much may be explored by means of variables determined at home. Given that categories of time use (e.g. time spent studying after school) are likely to be correlated with unobserved time-varying investments into children's learning (e.g. parental attention to schooling), coefficients on time use categories should be interpreted with care. The bounding exercise is also important in accounting for differences in enrolment across countries at different ages: as Table 1 documents, that nearly a third of the children in the Ethiopian sample are not yet enrolled at the age of 8 years while by the age of 15 drop-out rates are higher in India and Vietnam than in the other two countries.

Estimating achievement production functions is difficult as the full history of inputs applied at each age, as well as the full vector of child specific endowments, is not observed in any dataset. Specifications (3) and (4), which include previous test scores and a range of controls, are commonly known as ‘lagged value-added models’ (VAMs) of achievement production (Andrabi et. al., 2011; Todd and Wolpin, 2007). These models attempt to deal with this problem by entering the lagged achievement score in the estimation as a summary statistic for child-specific endowments and the full history of inputs. These VAMs will be used as the workhorse specifications for the analysis of divergence in this paper although I will investigate possibilities of bias due to unobserved heterogeneity and measurement error later in this section which, as Todd and Wolpin (2003, 2007) discuss, may importantly bias value-added (VA) estimates.²⁵

²⁵It is relevant to note here that a large recent literature finds the observed level of bias in VA estimates to be low across a range of applications including in comparisons with experimental estimates (Kane and Staiger, 2008; Kane, McCaffrey, Miller, and Staiger, 2013; Deming, Hastings, Kane, and Staiger, 2014; Angrist, Pathak, and Walters, 2013; Singh, 2013), with quasi-experimental estimates (Chetty. et. al., forthcoming), dynamic panel data estimates (Andrabi, Das, Khwaja, and Zajonc, 2011) and in simulated data with a variety of non-random assignment mechanisms (Guarino, Reckase and Wooldridge, forthcoming). My results in this paper, comparing VA estimates to those based on a fuzzy regression discontinuity design and unable to reject equality, further add to this literature.

Table 5: Descriptive statistics of control variables

	Ethiopia			India			Peru			Vietnam		
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
Panel A: Younger cohort												
<i>Child and background characteristics (X_{ic})</i>												
Male	0.53	0.5	1881	0.53	0.5	1903	0.5	0.5	1892	0.51	0.5	1916
First born	0.23	0.42	1881	0.39	0.49	1903	0.37	0.48	1892	0.46	0.5	1916
Caregiver's Education	2.95	3.73	1874	3.7	4.44	1900	7.75	4.64	1892	6.88	3.83	1908
Age in months	97.48	4.05	1879	96.03	3.92	1903	95.35	3.63	1890	97.09	3.75	1915
Height-for-age z-score	-1.21	1.05	1877	-1.44	1.03	1898	-1.14	1.03	1890	-1.07	1.05	1900
Wealth index (2006)	0.28	0.18	1881	0.46	0.2	1902	0.47	0.23	1892	0.51	0.2	1914
<i>Time use (hours spent on a typical day; $TU_{ic,a}$)</i>												
— Doing domestic tasks	1.66	1.37	1881	0.33	0.58	1903	0.87	0.7	1887	0.54	0.66	1899
— Tasks on family farm/business etc.	1.5	2.22	1880	0.01	0.1	1903	0.25	0.66	1886	0.09	0.48	1897
— Paid work outside household	0.01	0.28	1880	0.01	0.2	1903	0	0.08	1887	0	0.07	1897
— At school	4.91	2.54	1881	7.72	0.95	1903	6.02	0.9	1887	5.04	1.31	1898
— Studying outside school time	0.99	0.89	1881	1.86	1.09	1903	1.87	0.83	1886	2.82	1.49	1897
— General leisure etc.	4.44	2.39	1881	4.71	1.54	1903	4.13	1.65	1887	5.55	1.65	1898
— Caring for others	0.83	1.21	1881	0.21	0.5	1903	0.48	0.88	1886	0.24	0.66	1878
Panel B: Older cohort												
<i>Child and background characteristics (X_{ic})</i>												
Male	0.51	0.5	971	0.49	0.5	976	0.53	0.5	664	0.49	0.5	972
First born	0.2	0.4	971	0.31	0.46	976	0.31	0.46	664	0.37	0.48	972
Caregiver's Education	2.93	3.49	967	2.86	4.05	976	7.27	4.57	663	6.77	3.85	971
Age in months	180.34	3.58	971	179.76	4.24	975	179.1	4.1	661	181.12	3.83	972
Height-for-age z-score	-1.37	1.28	968	-1.64	1	970	-1.48	0.9	657	-1.43	0.91	967
Wealth index (2006)	0.3	0.17	971	0.47	0.2	976	0.52	0.23	664	0.52	0.19	970
<i>Time use (hours spent on a typical day; $TU_{ic,a}$)</i>												
— Doing domestic tasks	2.55	1.65	970	1.45	1.35	975	1.42	1.07	662	1.44	0.96	958
— Tasks on family farm/business etc.	1.34	2.09	970	0.49	1.72	975	0.68	1.49	662	1.05	2.13	958
— Paid work outside household	0.4	1.63	970	1.04	2.77	975	0.41	1.72	662	0.47	2	958
— At school	5.55	2.17	970	6.39	3.59	975	5.91	2.01	662	4.23	2.34	946
— Studying outside school time	1.84	1.23	970	2.01	1.54	975	2.09	1.12	662	3.06	2.13	941
— General leisure etc.	2.98	1.71	970	4.1	2.32	975	3.24	1.48	662	4.97	2.23	955
— Caring for others	0.67	0.93	970	0.28	0.75	975	0.73	1.18	662	0.16	0.64	951

Children in the older cohort were born in 1994-95 and children in the younger cohort in 2001-02. Caregiver's Education is defined in completed years; wealth index is an aggregate of various consumer durables and access to services at the household level. Height-for-age z-score is computed as per WHO standards. Unless indicated otherwise, the values of variables are from 2009.

Summary statistics of the controls used in the estimation of achievement production functions are presented in Table 5. Results from the estimation of Specifications 1-4 are presented for both cohorts in Table 6.

Table 6: Do home factors and child-specific endowments explain divergence?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
VARIABLES	8-years old (Younger cohort)				15-years old (Older cohort)			
<i>Country dummies</i>								
India	76.3*** (3.01)	64.5*** (2.92)	61.6*** (2.97)	16.3*** (3.58)	39.4*** (4.67)	13.8*** (4.11)	7.38* (4.29)	6.62 (4.23)
Peru	96.7*** (2.75)	79.1*** (2.71)	65.2*** (2.69)	48.2*** (2.85)	85.4*** (4.00)	62.6*** (3.53)	47.4*** (3.76)	57.6*** (3.81)
Vietnam	146*** (3.04)	127*** (3.06)	108*** (2.97)	92.2*** (3.45)	113*** (4.26)	66.0*** (3.95)	55.4*** (4.00)	69.5*** (4.26)
<i>Background characteristics</i>								
Male			3.79** (1.68)	4.94*** (1.60)			8.77*** (2.48)	11.1*** (2.47)
First-born child			6.56*** (1.68)	4.81*** (1.61)			4.38* (2.63)	2.61 (2.50)
Caregiver's education level			2.99*** (0.24)	2.18*** (0.23)			1.12*** (0.35)	0.33 (0.33)
Age in months			2.94*** (0.23)	2.80*** (0.22)			-0.51 (0.32)	0.033 (0.30)
Height-for-age z-score (2009)			10.3*** (1.05)	7.66*** (0.97)			5.49*** (1.35)	4.95*** (1.29)
Wealth index (2006)			74.6*** (5.36)	47.8*** (5.08)			67.1*** (7.70)	44.7*** (7.38)
<i>Time use (hours spent on a typical day)</i>								
— doing domestic tasks				0.76 (1.31)				2.70* (1.47)
— doing tasks on family farm etc.				-2.07* (1.22)				0.79 (1.30)
— doing paid work outside hh				0.32 (8.52)				0.29 (1.31)
— at school				13.6*** (0.99)				8.59*** (1.16)
— studying outside of school time				13.2*** (0.93)				8.07*** (1.19)
— general leisure etc.				1.64** (0.82)				3.11*** (1.09)
— caring for others				0.58 (1.25)				-0.60 (1.73)
Lagged test score (2006)		0.27*** (0.011)	0.12*** (0.0099)	0.10*** (0.0095)		0.48*** (0.016)	0.41*** (0.016)	0.33*** (0.014)
Constant	421*** (2.37)	301*** (5.51)	57.2** (22.5)	0.10 (23.8)	443*** (3.41)	228*** (8.32)	331*** (57.7)	194*** (57.0)
Observations	7,573	7,573	7,522	7,465	3,595	3,583	3,554	3,513
R-squared	0.285	0.352	0.450	0.514	0.197	0.397	0.441	0.504
<i>F-tests of equality of coefficients (p-value)</i>								
India = Peru	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
India=Vietnam	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Peru=Vietnam	0.00	0.00	0.00	0.00	0.00	0.28	0.00	0.00

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age.

In both cohorts, the mean differences in the test scores are statistically different across the four countries and substantial in magnitude (Cols. 1 and 5). Controlling for the lagged test achievement (Cols. 2 and 6) reduces the gap between countries somewhat in the younger cohort and substantially in the older cohort (eliminating most of the gap between Peru and Vietnam and between Ethiopia and India). Gaps decline further upon inclusion of background variables in X but the magnitude of decline is small as a proportion of the initial gap.

Inclusion of time use inputs has different effects in the two cohorts. In the older cohort aged 15, the gap between Ethiopia and Vietnam increases while in the younger cohort aged 8, the gap between Ethiopia and the other countries reduces substantially (especially with India where it is now at about a fifth of the initial cross-sectional gap). This latter pattern is likely a product of the enrolment profiles across country samples since a large proportion of 8-year old children in Ethiopia have not yet joined school, and because children in Vietnam and India are more likely to have left schooling by 15 than in the other two countries.

The central pattern in Table 6 is that a substantial gap remains between the country samples; even in the most extensive specifications, Ethiopia and Vietnam differ by between 0.7-0.9 SD at 8 and 15 years of age, accounting for more than 60% of the cross-sectional gap in test scores. These results suggest that while differences in endowments and socio-economic background play a role in creating differences across samples, it appears unlikely that this is the sole, or perhaps even the main, cause for divergence.

4.2 Does differential productivity of home inputs explain divergence?

Specifications 2-4 impose a strong assumption of common parameter coefficients on inputs across countries. This assumption is unlikely to hold; there is no reason to assume, for example, that a year of maternal education has an identical impact on child test scores in Vietnam and Ethiopia. In order to allow for maximum heterogeneity across the four countries, I estimate specifications (3) and (4) separately for each country thus allowing all input coefficients to differ. Results are presented in Tables 7 and 8 for the 8 year old and 15 year old sample respectively.

As can be seen, the coefficients on specific inputs differ across countries. It is, however, difficult to directly read from these Tables the importance of this differing productivity for explaining test score gaps. In order to facilitate such comparison, I present some counterfactual examples applying to each country sample, the input coefficients estimated in the different country samples i.e. predicting the mean level of achievement keeping the country's level of inputs fixed but varying the coefficients of the inputs (including the constant term) to match other countries.²⁶ The results are shown in Table 9 from specifications both with and without time use: the primary diagonal in each case, in bold, shows the actual level of achievement in the relevant country sample.

²⁶These are two polar cases where I change either all inputs or all coefficients. In practice, from a policy perspective, it may not be feasible or even desirable to change all inputs or coefficients nor is choice limited to only choosing to shift elements of only one or the other vector; many more combinations could be explored. The purpose here is only to highlight two contrasting possibilities to assess the relative importance of these two channels (differences in the level of inputs and differences in input productivity) in explaining divergence.

Table 7: Country-specific production functions of achievement: 8 years old

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Without time use				With time use			
VARIABLES	Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Male	1.33 (6.55)	5.04 (3.19)	8.95*** (2.47)	-0.55 (2.65)	2.07 (5.36)	4.52 (3.52)	9.32*** (2.76)	-0.25 (2.66)
Eldest	7.17 (4.54)	3.59 (3.05)	9.37*** (3.02)	7.74* (3.77)	3.38 (3.66)	2.75 (3.35)	7.89** (3.30)	8.26* (4.06)
Caregiver's education level	3.37*** (0.79)	2.91*** (0.82)	2.69*** (0.46)	3.73*** (0.88)	2.28*** (0.56)	2.25*** (0.58)	2.52*** (0.43)	2.34*** (0.81)
Age in months	2.73*** (0.52)	1.96*** (0.60)	2.67*** (0.35)	4.07*** (0.54)	2.23*** (0.50)	1.95*** (0.56)	2.67*** (0.33)	4.16*** (0.56)
Height-for-age (2009)	14.4*** (2.66)	9.93*** (2.07)	6.77*** (2.09)	10.3*** (3.12)	7.55*** (2.33)	8.88*** (1.96)	6.27*** (1.89)	7.33*** (2.17)
Wealth index (2006)	174*** (27.7)	44.7 (26.6)	25.3*** (7.86)	91.7*** (27.5)	106*** (18.4)	21.5 (19.7)	25.2*** (7.99)	64.9*** (20.2)
Time use (hours on a typical day)								
— doing domestic tasks					0.47 (3.66)	3.64 (4.37)	7.37*** (2.06)	-4.15 (4.46)
— doing tasks on family farm etc.					0.67 (3.55)	-16.5*** (5.58)	-0.44 (1.95)	-24.7*** (4.86)
— doing paid work outside hh					-4.85 (8.81)	22.0*** (6.56)	-5.49 (4.95)	15.5 (9.32)
— at school					12.6*** (3.57)	22.4*** (2.64)	9.21*** (2.98)	5.16 (4.84)
— studying outside school time					19.8*** (3.76)	19.9*** (5.03)	7.61*** (1.52)	4.65 (3.42)
— general leisure etc.					1.26 (3.20)	5.36* (2.86)	2.33* (1.19)	-2.71 (2.98)
— caring for others					2.11 (4.71)	0.96 (5.04)	2.00 (1.17)	-8.75 (6.38)
Lagged CDA score (2006)	0.071** (0.026)	0.15*** (0.029)	0.13*** (0.020)	0.12*** (0.040)	0.044* (0.023)	0.14*** (0.028)	0.13*** (0.019)	0.089** (0.031)
Constant	77.9 (47.6)	212*** (59.5)	162*** (31.5)	47.4 (55.8)	61.7 (63.5)	-6.91 (74.0)	76.7** (32.8)	56.2 (61.8)
Observations	1,835	1,892	1,888	1,907	1,834	1,892	1,881	1,858
R-squared	0.255	0.177	0.282	0.309	0.374	0.280	0.312	0.353

Robust standard errors in parentheses. Standard errors are clustered at site level. *** p<0.01, ** p<0.05, * p<0.1

Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age.

Table 8: Country-specific production functions of achievement: 15 years old

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Without time use				With time use			
	Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Male	24.0*** (7.87)	22.0*** (5.71)	-6.11 (4.10)	-7.49 (4.37)	27.8*** (7.42)	19.8*** (6.53)	-0.90 (3.71)	-3.05 (4.19)
Eldest	0.46 (9.40)	8.29* (4.16)	6.53* (3.37)	3.03 (3.60)	-1.75 (8.78)	6.88 (4.11)	6.79* (3.37)	-0.065 (3.73)
Caregiver's education level	-1.36* (0.77)	2.22*** (0.73)	1.09*** (0.34)	2.88*** (0.96)	-1.33 (0.80)	0.79 (0.62)	0.80*** (0.28)	2.17** (0.91)
Age in months	-0.34 (0.95)	-0.86 (0.63)	0.46 (0.40)	-0.77 (0.85)	-0.12 (0.93)	-0.0082 (0.60)	0.52 (0.35)	-0.017 (0.85)
Height-for-age (2009)	8.32** (3.07)	3.72 (3.03)	4.61* (2.55)	4.97* (2.77)	6.73** (2.78)	3.39 (2.88)	4.80* (2.46)	4.55 (2.97)
Wealth index (2006)	120*** (29.1)	46.5** (19.5)	34.1*** (8.76)	85.1*** (24.9)	88.0*** (28.4)	28.1 (17.1)	22.5*** (7.75)	66.0** (24.0)
Time use (hours on a typical day)								
— doing domestic tasks					6.14* (3.05)	5.37 (3.42)	3.44* (1.81)	2.79 (3.77)
— doing tasks on family farm etc.					3.30 (3.72)	-3.82 (2.61)	0.46 (1.98)	-0.71 (2.44)
— doing paid work outside hh					4.02 (3.84)	-4.70** (2.20)	2.23 (2.40)	0.78 (1.63)
— at school					10.7*** (2.89)	5.46** (2.42)	7.35*** (2.28)	6.33** (3.00)
— studying outside school time					17.6*** (2.83)	7.99*** (2.09)	5.38*** (1.53)	1.32 (1.90)
— general leisure etc.					7.11* (3.64)	1.03 (1.66)	1.40 (1.37)	-1.29 (1.97)
— caring for others					5.41 (3.84)	-9.77** (3.69)	1.74 (1.06)	-1.59 (4.08)
Lagged math score (2006)	0.64*** (0.048)	0.49*** (0.059)	0.22*** (0.040)	0.26*** (0.034)	0.55*** (0.041)	0.37*** (0.042)	0.19*** (0.032)	0.22*** (0.032)
Constant	184 (177)	357*** (114)	321*** (74.8)	499*** (142)	50.2 (176)	223* (123)	262*** (64.5)	369** (140)
Observations	964	970	656	964	963	969	656	925
R-squared	0.317	0.369	0.304	0.348	0.369	0.495	0.390	0.397

Robust standard errors in parentheses. Standard errors are clustered at site level.*** p<0.01, ** p<0.05, * p<0.1
Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age.

The results are informative and telling. In the younger (8 year old) sample, the difference between the average levels of achievement between Ethiopia and India seems mostly a difference in the inputs of the children (including their test scores at age 5 which reflect investments in early childhood): equalizing these in the specifications with time use, but maintaining the same production function parameters as estimated in Table

7 for the country, reduces almost the entire gap between Ethiopia and India.²⁷ However, strikingly, the difference between Vietnam, the only ‘high performer’ in our sample, and the other three countries seems to lie not in the endowments, including what children had learnt prior to school entry, but in the higher rates of learning afterwards: for each of the other three countries, considerably more of the learning gap is closed by equalizing productivity of inputs than by equalizing the level of inputs. For example, both in specifications with and without time use, while raising Peruvian inputs to Vietnamese levels does not change mean achievement levels in Peru at all, equalizing the productivity of the inputs closes almost the entire gap; similarly, in the specifications with time use, the whole gap between India and Vietnam is covered by equalizing the productivity of inputs to Vietnam.²⁸

²⁷To see this, note that fixing inputs to Indian levels produces a predicted mean level of 487 using Ethiopian coefficients and 497 using Indian coefficients. The difference between the two numbers is not statistically significant and we cannot reject that the full gap is thus closed. The reliance on the specifications incorporating time use is particularly relevant here since it captures the differences in enrolment between Ethiopia and the other countries at 8 years of age.

²⁸To see this note that, in the specifications with time use variables, applying Vietnamese coefficients to Peruvian inputs shifts the Peruvian mean score from 518 to 558 but shifting inputs (while keeping coefficients unchanged) does not move scores at all. Similarly, applying Vietnamese coefficients to Indian inputs raises Indian mean scores from 497 to 563 which is near-identical to the Vietnamese mean; the reverse case (keeping Indian coefficients but applying Vietnamese inputs) reduces achievement, albeit not significantly.

Table 9: Predicted mean achievement levels under various counterfactual scenarios

Younger cohort (8-years)									
Coefficients (β_c)									
Without time use					With time use				
		Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Inputs ($X_{ic}; TU_{ica}$) $Y_{ic,a-1}$	Ethiopia	420.79 (9.87)	485.28 (10.64)	495.47 (5.49)	523.15 (13.48)	420.75 (10.85)	390.94 (16.72)	486.66 (9.62)	488.38 (19.19)
	India	450.36 (11.54)	497.32 (9.59)	503.74 (4.97)	539.9 (11.02)	487.38 (10.39)	497.32 (9.87)	516.86 (7.99)	563.24 (14.79)
	Peru	470.66 (11.35)	514.64 (10.7)	517.73 (4.65)	559.32 (10.53)	479.48 (10.93)	468.87 (10.96)	517.74 (5.65)	557.66 (11.68)
	Vietnam	478.69 (11.08)	518.05 (9.76)	522.35 (4.51)	567.03 (9.16)	492.1 (12.06)	476.78 (13.14)	520.84 (7.09)	568.22 (11.43)
Older cohort (15-years)									
Coefficients (β_c)									
Without time use					With time use				
		Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Inputs ($X_{ic}; TU_{ica}$) $Y_{ic,a-1}$	Ethiopia	443.17 (10.54)	448.98 (10.14)	507.7 (7.12)	502.26 (9.21)	443.15 (12.13)	453.52 (11.38)	512.33 (8.5)	524.55 (12.03)
	India	495.01 (13.12)	482.61 (9.84)	524.44 (6.54)	529.91 (8.67)	496.15 (14.96)	482.86 (10.38)	531.03 (9.05)	549.28 (12.55)
	Peru	493.1 (12.65)	493.53 (9.86)	529.74 (6.04)	546.1 (8.74)	483.25 (14.14)	481.28 (10.68)	529.74 (7.25)	557.88 (11.58)
	Vietnam	525.34 (12.65)	515.18 (10.25)	542.1 (6.56)	557.05 (8.54)	521.01 (14.1)	504.53 (11.36)	535.76 (9.14)	558.18 (10.56)

Cells contain linear predictions of test scores using combinations of country-specific production function parameters (β_c), as estimated in Tables 7 and 8 for 8-year and 15-year olds respectively with country-specific input levels (X_{ic} and TU_{ic}) as in Table 5. Each row shows predicted values of mean achievement when applying, to a given country sample, different country-specific coefficients indicated in column headings. Results are shown for specifications with and without time use categories. Standard errors of predictions in parentheses.

It is clear from this exercise is that, whereas child endowments and the investments made in early childhood are undeniably important, the major divergence with Vietnam is after the age of 5 years. Considerably less clear is the source of this divergence. The major difference in productivity

between Vietnam and the rest comes from the difference in the coefficients on the age in months in Table 7: we know that Vietnamese children seem to be learning more as they age each month than the children in the other countries but we don't quite know why.²⁹ This will be investigated in the later subsections.

Results for the older cohort are rather more mixed. Similar to the result at 8 years, the gap at 15 between Ethiopia and India seems to be entirely a product of the differences in the endowments across samples including the amount learnt till the age of 12 years. In closing the gaps between Vietnam and the other countries, the picture is not quite as clear as at 8 years: with Ethiopia, in specifications with time use, equalizing the level of all inputs seems as effective as equalizing all production function parameters; while the same is suggested with Peru in specifications without time use, once time use is incorporated this implies a slightly larger gain from equalizing coefficients; with India, in both sets of results, equalizing coefficients seems more effective in closing the gap with Vietnam compared to equalizing inputs.

4.3 Do differential exposure to schooling and differential productivity of schooling explain divergence?

As Table 9 documents, only a small portion of the divergence across countries (especially with Vietnam) till the age of 8 years is accounted for by the levels and differential productivity of home inputs across the

²⁹The difference between the coefficient on age in Vietnam and in the other countries is invariably statistically significant in cross-equation tests.

four country samples. One possibility that may account for divergence after 5 is the differential exposure to schooling across the four country samples; for example, Ethiopian children enter school much later than in the other countries (Table 1) and thus have less schooling at every age in the sample than the other countries. Similarly we would expect given previous work (Hanushek and Kimko, 2000; ?; Schoellman, 2012) that the quality of schooling differs across these contexts, which could also contribute to the growth of these gaps.

In order to study the importance of these schooling-based sources of divergence, I estimate the following specifications:

$$Y_{ic,a} = \alpha_c + \beta_1.Y_{ic,a-1} + \beta_2.X_{ic} + \beta_3.grade_{ica} \quad (5)$$

$$+ \beta_4.TU_{ic,a} + \epsilon_{ica} \quad (6)$$

where in addition to variables defined previously, I also include a variable for the highest grade completed by the child at age a . As in the previous specification, the estimation is carried out separately for each country sample and I estimate the production function both with and without the time use inputs. The parameter of interest is β_3 which, if it differs across countries, would indicate differences in the amount of progress in quantitative skills per grade completed across the different educational systems.³⁰

³⁰Note that grades completed may be regarded as an *outcome* of educational systems rather than merely an input into learning and thus raise concerns about its endogeneity (for example if, as is likely, the same factors determine both grade completion and amount learnt in school). Identification of β_3 in this case rests on the assumption that all such factors are either directly controlled for in the estimation or effectively proxied

Table 10: Comparing effectiveness of a grade of schooling: 8-years old

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Without time use				With time use			
	Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Highest grade completed	40.9*** (4.67)	27.4*** (2.03)	33.6*** (3.60)	60.9*** (14.6)	28.4*** (4.48)	25.4*** (1.62)	32.6*** (3.55)	55.2*** (10.9)
Male	3.26 (5.61)	12.7*** (3.05)	8.73*** (2.22)	1.65 (2.39)	4.44 (4.82)	11.6*** (3.13)	8.92*** (2.47)	1.62 (2.66)
Eldest	3.97 (4.08)	5.74** (2.60)	8.45*** (2.89)	6.63** (2.98)	1.63 (3.72)	4.59 (3.01)	7.09** (3.09)	7.18** (3.16)
Caregiver's education level	3.76*** (0.66)	2.40*** (0.70)	2.23*** (0.49)	3.16*** (0.80)	2.74*** (0.52)	1.86*** (0.49)	2.10*** (0.48)	2.18*** (0.72)
Age in months	1.26** (0.53)	0.51 (0.45)	-0.067 (0.30)	0.18 (1.10)	1.30** (0.56)	0.61 (0.41)	0.0079 (0.30)	0.69 (0.87)
Height-for-age (2009)	9.31*** (2.64)	5.38** (2.21)	5.22** (1.92)	7.14*** (1.78)	5.30** (2.33)	4.79** (1.85)	4.82** (1.73)	4.81*** (1.56)
Wealth index (2006)	151*** (25.9)	53.6** (23.8)	17.6* (8.80)	78.3*** (20.9)	105*** (18.8)	31.0* (17.8)	18.1* (8.91)	59.0*** (19.0)
Time use (hours on a typical day)								
— doing domestic tasks					2.23 (3.37)	3.06 (4.21)	6.72*** (1.95)	-4.16 (3.89)
— doing tasks on family farm etc.					1.41 (3.44)	-13.6*** (3.39)	0.12 (1.61)	-21.4*** (5.34)
— doing paid work outside hh					-3.94 (7.86)	22.3*** (7.50)	-4.05 (3.84)	-2.56 (7.02)
— at school					12.1*** (3.36)	21.2*** (2.54)	8.94*** (2.90)	3.78 (4.09)
— studying outside school time					13.8*** (3.72)	17.7*** (4.88)	6.77*** (1.71)	2.03 (3.09)
— general leisure etc.					2.12 (3.24)	4.53* (2.52)	2.40* (1.31)	-2.64 (2.59)
— caring for others					3.29 (4.65)	1.74 (4.70)	1.96* (1.05)	-7.01 (4.77)
Lagged CDA scores (2006)	0.067*** (0.023)	0.13*** (0.027)	0.100*** (0.021)	0.065* (0.032)	0.045* (0.022)	0.12*** (0.027)	0.100*** (0.020)	0.049 (0.030)
Constant	196*** (49.2)	306*** (45.5)	401*** (29.5)	354*** (74.1)	129* (72.0)	97.6* (53.8)	313*** (38.8)	333*** (65.5)
Observations	1,835	1,892	1,888	1,907	1,834	1,892	1,881	1,858
R-squared	0.340	0.276	0.343	0.437	0.410	0.365	0.370	0.458

Robust standard errors in parentheses. Standard errors are clustered at site level. *** p<0.01, ** p<0.05, * p<0.1
Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age.

I present the estimated production function estimates for an additional grade completed in each country in Table 10 for the younger cohort.

for by the lagged achievement score, which is the maintained assumption underlying value-added models. In Sec. 4.4 I will document how this channel of potential bias does not seem to be important in the case of Peru and Vietnam, where I am able to generate alternative IV estimates.

The results for younger children are striking: the learning increment per additional grade completed is much larger in the Vietnamese sample than in the Indian or Peruvian samples, a conclusion that is unchanged whether or not time use categories are included. These differences are statistically significant and the learning increment per year in Vietnam is significantly greater than the increment in any of the other countries. The difference with Peru is particularly striking since these two samples had similar baseline achievement levels at 5.

Does incorporating the differential effectiveness of schooling in the four country samples enable us to account for a larger proportion of the divergence between countries? The important pattern to note is that the inclusion of grades completed has removed the higher maturation effect (coefficient on age) in Vietnam in comparison to other countries: the pattern noted earlier, that Vietnamese children seemed to be learning more than in the other countries, disappears upon including grades completed and allowing for differential effectiveness. The contribution of differential effectiveness of grades completed to the divergence in test scores is large: for example, raising the effectiveness of a grade of schooling to Vietnamese levels, even keeping all endowments (including learning at 5) as well as all other coefficients unchanged, closes the gap between India and Vietnam by about 70% and between Peru and Vietnam almost entirely.³¹

³¹To see this, note that raising Indian productivity to Vietnamese levels would lead to an additional about 30 points per grade (the difference between the coefficient on grade in Col. 8 and Col. 6) which when multiplied by the mean number of grades completed (1.63 years, Table 1) closes the gap by about 49 points; this is just above 70% of the mean gap between the two countries of 70 points. The Ethiopian case is somewhat different from Peru and India since both the level and productivity of grades completed are lower than Vietnam. Here also, if the sample had the same amount of schooling and grade productivity as Vietnam, it would close about 60% of the gap between the two countries.

Table 11: Comparing effectiveness of a grade of schooling: 15-years old

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Without time use				With time use			
	Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Highest grade completed	21.7*** (2.14)	19.9*** (1.72)	17.8*** (3.48)	12.2*** (2.67)	19.1*** (2.31)	13.1*** (1.37)	15.1*** (3.09)	10.3*** (2.77)
Male	25.8*** (6.96)	22.0*** (4.57)	-1.57 (3.59)	-5.77 (4.34)	26.9*** (6.02)	18.8*** (6.08)	0.75 (3.69)	-2.59 (4.07)
Eldest	-0.74 (7.91)	3.38 (3.68)	3.62 (2.70)	3.19 (3.60)	-1.62 (7.50)	4.24 (3.95)	4.24 (2.92)	0.67 (3.61)
Caregiver's education level	-1.32* (0.71)	2.01*** (0.55)	0.62* (0.34)	2.43** (0.91)	-1.30* (0.72)	1.02* (0.54)	0.51 (0.29)	1.88** (0.86)
Age in months	-0.49 (0.90)	-1.30** (0.53)	-0.98* (0.47)	-1.02 (0.82)	-0.36 (0.89)	-0.57 (0.56)	-0.74* (0.39)	-0.35 (0.83)
Height-for-age (2009)	1.11 (2.89)	-0.61 (2.81)	1.90 (2.10)	3.48 (2.79)	1.09 (2.76)	0.54 (2.78)	2.23 (2.01)	3.47 (2.88)
Wealth index (2006)	76.8*** (24.7)	34.2* (18.7)	6.09 (8.04)	68.2*** (23.7)	65.2** (23.9)	25.2 (16.8)	3.85 (7.88)	56.5** (23.3)
<i>Time use (hours on a typical day)</i>								
— doing domestic tasks					4.30 (3.21)	2.88 (2.80)	2.52 (1.61)	1.39 (3.72)
— doing tasks on family farm etc.					4.07 (3.35)	-2.43 (2.41)	0.61 (1.60)	-0.57 (2.58)
— doing paid work outside hh					4.09 (3.31)	-3.29 (2.05)	2.49 (1.90)	-0.19 (1.67)
— at school					8.16** (3.06)	3.38 (2.31)	5.29*** (1.66)	4.84 (2.94)
— studying outside school time					10.5*** (2.87)	6.08*** (2.01)	3.13** (1.16)	0.60 (2.08)
— general leisure etc.					5.52 (3.63)	-0.067 (1.70)	1.65 (1.19)	-1.46 (2.10)
— caring for others					7.13* (3.63)	-9.65** (4.07)	1.16 (0.96)	-2.86 (4.05)
Lagged math scores (2006)	0.34*** (0.046)	0.34*** (0.051)	0.14*** (0.019)	0.23*** (0.029)	0.33*** (0.046)	0.31*** (0.041)	0.14*** (0.016)	0.20*** (0.030)
Constant	227 (167)	349*** (96.2)	491*** (73.7)	474*** (133)	122 (163)	267** (114)	424*** (55.8)	375** (135)
Observations	964	970	656	964	963	969	656	925
R-squared	0.427	0.482	0.454	0.408	0.443	0.533	0.486	0.435

Robust standard errors in parentheses. Standard errors are clustered at site level. *** p<0.01, ** p<0.05, * p<0.1
Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age.

Results for the older cohort (Table 11) are mixed and stand in contrast to the results at 8 years of age: in particular, the progress in Vietnam appears lower than in other countries although, in specifications which include time use categories, the coefficient on highest grade schooling in Vietnam is not significantly different from India or Peru. This is surprising and contrasts

with all the previous patterns highlighted in the data. A likely possibility which could account for this anomalous finding is that as children move to secondary school, the focus of math training moves away from arithmetic and basic geometry (the focus of the testing in Young Lives) to more challenging topics such as trigonometry, probability and calculus which are not assessed here; if such a movement is more pronounced in Vietnam than elsewhere, or if Vietnamese children had already substantially mastered the range of skills tested by the age of 12, it is possible that we do not see much progress at all on the available test metric.³²

The important point to stress regarding the divergence between 12 to 15 years is that the differential productivity of schooling, which is the most salient contemporaneous policy variable among the inputs in the production function, does not account for the divergence between country samples.

4.4 Are VA estimates reliable? Comparison with IV estimates

Value-added models are based on the identifying assumption that the lagged test score suffices to proxy for any relevant sources of bias in the interpretation of input coefficients in the production function estimates. Such concerns, relating to possible endogeneity of inputs, are particularly salient for grades completed in school. Within-country variation in this

³²Note that the coefficient in Ethiopia is invariably larger than the coefficient in the other countries both with and without time use; this difference is statistically significant from Vietnam in both specifications and from India in the specification with time use. Combined with the fact that the median grade completed for this age group in the Ethiopian sample is Grade 6 while in the Indian and Vietnamese samples it is Grade 9, this could be indicative that the skills tested by the Young Lives assessment are mostly produced at lower grade levels.

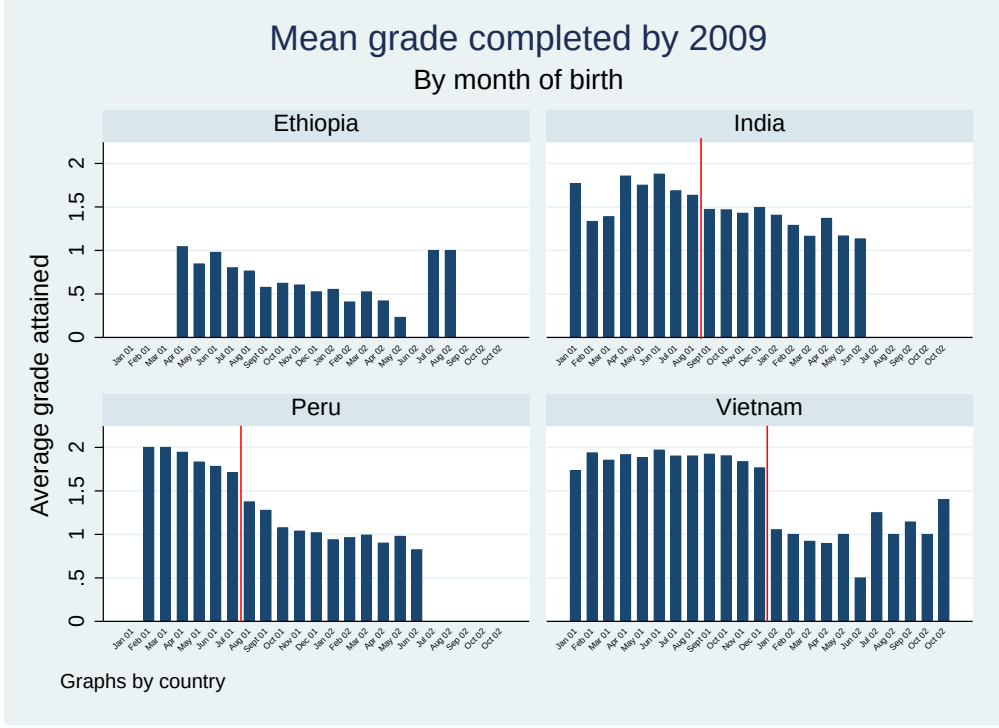
variable comes from three possible sources: the age of starting school, retention in particular grades due to lack of academic progress, and early or intermittent dropping out. The importance of these sources differs across the educational trajectory: whereas differences in the age of starting school account for the bulk of the variation at younger ages, by the age of 15 grade repetition and drop-out (both of which may plausibly be caused by low academic achievement) are both more relevant. If the factors that determine these three channels are effectively proxied by lagged achievement, the estimates can be interpreted causally but not otherwise. In this section, I estimate causal impacts of a grade of schooling based on plausibly exogenous variation arising from enrolment guidelines and assess if any conclusions are substantively changed.

My strategy for estimating causal effects of additional grades completed uses variation in when children joined school, arising from their month of birth and the enrolment guidelines of particular countries, to instrument grades completed in the VA specifications presented in Table 9. Figure 4 presents the average number of grades completed by children born in different months in the sample in each country. As can be seen there is a discontinuity in Vietnam between Dec 2001-Jan 2002 and a somewhat fuzzier discontinuity in Peru between Jul-Aug 2001; these also represent the official guidelines for enrolment of children into first grade in these two countries (highlighted by red reference lines).³³

³³Guidelines for enrolment in first grade in Peru in the 2007 academic year state that children should have completed 6 years of age by July 31, 2007, thus generating the discontinuity. In Vietnam, guidelines stipulate that the child should be enrolled in school in the calendar year that he/she turns six years of age, thus generating a discontinuity in grades completed between children born in December and January.

While there are similar guidelines in India as well, requiring in Andhra Pradesh all children to have turned 5 by Sept. 1 of the year in which admission is sought, the discontinuity created is much less sharp and seems inadequate in statistical power to

Figure 4: Discontinuity in grade attained by month of birth



The instrumentation strategy implies a first-stage equation of the form:

$$grades_{i,2009} = \mu + \gamma_1.Threshold_i + \gamma_2.X_i + \gamma_3.site_i + \epsilon \quad (7)$$

where Threshold is defined as an indicator variable equalling 1 if born after July 2001 in Peru or after Dec 2001 in Vietnam and 0 otherwise. X is the vector of controls listed in previous specifications and includes the child's age in months.

be used as an IV by this point of the children's trajectory. Using this discontinuity, I obtain very imprecise estimates, which are not statistically distinguishable from the OLS (VA) estimates, from zero, or from the coefficients of any of the other countries, thus not allowing for any firm conclusions to be drawn; a similar conclusion is also borne by using Dec 01-Jan 02 as the relevant threshold, as used by Singh, Park, and Dercon (2014).

There is also no evidence of such discontinuities that can be used in Ethiopia.

Table 12: Discontinuity-based results on grade effectiveness in Peru and Vietnam

VARIABLES	(1)	(2)	(3)	(4)
	Dep var: Math scores (2009)			
	Peru			Vietnam
Highest grade completed	20.1*** (7.61)	20.9*** (7.96)	47.3*** (7.49)	46.3*** (7.16)
Male	9.43*** (2.39)	9.96*** (2.63)	1.34 (2.36)	1.56 (2.46)
Eldest	8.04*** (2.74)	6.44** (3.01)	5.14 (3.16)	6.36** (3.05)
Caregiver's education level	2.31*** (0.40)	2.14*** (0.37)	3.05*** (0.61)	2.41*** (0.55)
Age in months	0.94 (0.66)	0.87 (0.71)	0.41 (0.57)	0.64 (0.53)
Height-for-age (2009)	6.15*** (2.20)	5.59*** (2.00)	6.00*** (1.96)	4.18*** (1.44)
Wealth index (2006)	29.7*** (7.67)	29.0*** (7.84)	40.2** (16.2)	28.6** (13.4)
<i>Time use (hours on a typical day)</i>				
— doing domestic tasks		4.99*** (1.71)		-3.37 (4.40)
— doing tasks on family farm etc.		-0.13 (2.20)		-15.1*** (4.86)
— doing paid work outside hh		0.42 (4.20)		-2.32 (6.37)
— at school		8.45*** (2.88)		7.79 (4.89)
— studying outside school time		6.72*** (1.54)		9.54*** (2.73)
— general leisure etc.		0.95 (1.16)		0.18 (1.77)
— caring for others		2.08** (0.96)		-4.37 (4.07)
Lagged math scores (2006)	0.13*** (0.020)	0.12*** (0.020)	0.11*** (0.031)	0.088*** (0.027)
Constant	290*** (58.2)	227*** (69.2)	375*** (55.5)	316*** (60.2)
Observations	1,888	1,881	1,907	1,858
R-squared	0.366	0.393	0.481	0.504
Kleibergen-Paap F-statistic	108	110	113	152

Robust standard errors in parentheses. Standard errors are clustered at site level. *** p<0.01, ** p<0.05, * p<0.1
Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age. Estimation includes a vector of site fixed effects, coefficients for which are not reported.
Highest grade completed is treated as endogenous in this table and instrumented for using in each country a discontinuity arising from en

The second stage, instrumenting grades completed with $Threshold_i$, is given by the following equation which is identical to Eqs. (5) and (6) but for additionally including site fixed effects ($site_i$) within country in order

to absorb any differences across sites in the implementation of enrolment guidelines.³⁴

$$Y_{ic,a} = \alpha_c + \beta_1.Y_{ic,a-1} + \beta_2.X_{ic} + \beta_3.grade_{ica} + \gamma.site_i + \epsilon_{ica} \quad (8)$$

$$+ \beta_4.TU_{ic,a} \quad (9)$$

Results from the estimation are presented in Table 12. As can be seen, in both Peru and Vietnam, the coefficients on grades completed are similar to, if somewhat than, the coefficients obtained from the OLS VA models in Table 9 and coefficients on most other variables are also unchanged.³⁵ In both countries, the OLS VA coefficients lie within the 95% confidence intervals of the IV estimates. More pertinently from our perspective, the differences indicated between the productivity of a year of schooling in Peru and Vietnam are also unchanged. In short, the VA models do not appear to be biased in these two samples.³⁶

³⁴The inclusion of site fixed effects is appropriate in this setting since we are not comparing the constant terms across countries (unlike in previous subsections). It is useful to note, however, that it does not notably alter our conclusions even if site fixed effects are excluded from the IV specifications: the core result, of the coefficient in Vietnam being considerably higher than in Peru is unchanged. One important difference is that in the absence of site fixed effects, the coefficient on a year of schooling in Peru is no longer statistically distinguishable from zero.

³⁵In interpreting the IV estimates, it should be remembered that these are Local Average Treatment Effects identified over the compliers who are prompted to join school as a result of the discontinuity. If the effects of grade effectiveness are heterogeneous, with the youngest children in class gaining less than their older peers, a decline in the coefficient does not necessarily indicate bias. This is, however, a point of marginal concern in this particular instance since the OLS VA estimates and the IV estimates are not statistically different and the pattern across countries is entirely unchanged.

³⁶While this is not direct evidence supporting the validity of VA estimates in Ethiopia and India, it is suggestive that the production function estimates are reliable. Unfortunately, these discontinuities seem to lack explanatory power in the first-stage at the age of 15 years. So I cannot similarly test the validity of the VA estimates for the older cohort.

4.5 Robustness checks

Flexible lag structure

In the analysis thus far dynamics have been modelled linearly with the lagged achievement measure entering the regression specifications in levels. If growth trajectories of achievement are in fact non-linear, value-added estimates may suffer from a misspecification bias. In order to test for this possibility, I re-estimated the production function using a third-order polynomial of the lagged test score instead of the lag in levels, following the practice in Chetty, Friedman, and Rockoff (2014).

Results from this exercise are reported in Appendix B. In the 8-year sample, parameters on coefficients on background covariates, time use and grades completed are not significantly altered, indicating that our key results are unchanged. In the 15-year old sample, there are some differences in the coefficient on grades of schooling which declines in India and Peru but the main comparative conclusions about rates of progress across countries (or the contribution of different factors to divergence between 12 and 15 years of age) are not changed.

Bias due to censoring in the lagged achievement score for 15-year olds

As noted in Section 2 and 3, there are non-trivial ceiling and floor effects in the 12-year old maths test, leading to a censoring of the true achievement distribution. Recent work by Rigobon and Stoker (2007, 2009) demonstrates that such censoring in regressors can cause bias in estimates; given that the severity of the censoring varies across the four country

samples, it is possible that different degrees of bias could exist and affect the comparative results in this paper for 15-year olds.

To test the empirical relevance of this bias, I use Bayesian *expected a posteriori* (EAP) IRT scores from the 12-year old assessment instead of maximum likelihood estimates of ability as the lagged achievement measure and rerun the estimation reported in Table 11. EAP scores utilize information from a prior distribution of ability along with the log likelihood and are less affected by censoring issues.³⁷ The resulting estimates are presented in Appendix C. As may be noted, although the coefficients on the lagged achievement measure rise markedly in both Vietnam and India (which were most affected by censoring) most input parameters, including most importantly the coefficient on grade completed, are unchanged indicating that our conclusions continue to hold. Notably, the coefficient on lagged achievement looks considerably more similar across countries correcting for the censoring than in Tables 8 and 11.

Measurement error in lagged achievement

Test scores are noisy measures of (latent) academic knowledge. This measurement error in lagged achievement can cause bias in the estimated production function parameters. In order to test for this possibility in the younger cohort, I instrumented the lagged quantitative achievement measure (CDA at age 5) with the scores of the child in a test of receptive vocabulary that was taken at the same time.³⁸ Informativeness of the IV

³⁷Please see Das and Zajonc (2010) for an explanation of technical details of Bayesian EAP estimation of IRT models.

³⁸Since the vocabulary tests are administered in different languages, with corresponding differences in difficulty, I cannot directly compare them across countries.

rests on the correlation between different domains of cognitive achievement and first-stage results are strong. The validity of the IV rests on the assumption that the measurement error in the two tests, conducted at the same time, is independent of each other.³⁹

Results are reported in Appendix D. In this sample, coefficients on the grade of schooling and other inputs are not materially affected. Coefficients on the lagged achievement measure rise significantly in Ethiopia and Peru, consistent with attenuation bias due to measurement error in lagged achievement, but remain statistically indistinguishable from OLS results in India and Vietnam. The substantive results regarding the differential effectiveness of a grade of schooling in different countries and its contribution to divergence in achievement are not affected.

Overall, the range of robustness checks indicate that while there may be uncertainty regarding the ‘true’ value of the persistence parameter due to measurement error in the lagged measure, and the functional form in which it enters the estimation, it does not seem to change the main conclusions that are drawn in the paper from the various specifications: there is divergence in both age groups even conditional on test scores and it is explained by differential grade productivity in primary school between 5-8 years of age but not in the older cohort between 12-15 years.

However, I can use them as instruments, utilizing the (within-country) correlation between math and vocabulary scores.

³⁹This is a strong assumption which rules out correlated shocks between different test outcomes, for example measurement error due to testing conditions on the day of assessment, but is often used in this literature to correct for measurement error. This exclusion restriction is not maintainable in the older cohort due to the censoring of lagged achievement. Since the censoring of lagged achievement necessarily implies that measurement error is correlated with ability, an independent measure of ability (such as the vocabulary test) cannot be a valid IV for the lagged achievement.

5 Conclusions

In this paper I have characterized the emergence and evolution of test score gaps in quantitative ability across using panel data on two cohorts of children in Ethiopia, India, Peru and Vietnam. Furthermore, I have decomposed the divergence of test scores, between 5-8 years of age and between 12-15 years of age, into various proximate sources.

Several results stand out. Achievement levels in three of the four countries are very low by international standards. Gaps between countries open up early and show evidence of increasing over the educational trajectory, thus preserving the ordering of country samples apparent at the age of 5 years. Estimates from value-added models indicate that this divergence is not wholly (or even mostly) accounted for by differences in child-specific individual and home endowments, although results suggest that differences in early investments (embodied in test scores prior to school entry) have long-lasting effects: there is significant difference in the exposure to and effectiveness of schooling in the four samples which accounts for an important portion of the gap, especially at primary level. Results from VAMs seem unbiased based on comparison with IV estimates for 8-year old children in Peru and Vietnam.

It is important that the differences in the levels of test scores at 8 years are accounted for almost entirely by the difference in the quality of schooling across countries: this matters particularly because school-based learning may be easier for policy to directly influence than (potentially unobserved) investments into children at home. This also provides a clear link between this paper and the vast literature on such interventions in developing

countries (Glewwe and Kremer, 2006; Kremer, Brannen, and Glennerster, 2013; McEwan, 2013); while I document that school productivity is an important source of divergence and thereby imply the need for educational interventions, the impact evaluation literature rigorously identifies the tools by which learning gains per year may be enhanced in a variety of contexts. The persistence of gaps between the ages of 12 and 15 seems likely to be a product of investments made prior to 12 years i.e investments in early childhood and primary school.⁴⁰

The analysis attempts also to show how linked panel data across countries could greatly aid the understanding of learning gaps across countries. Even though international testing programs like PISA and TIMSS are steadily increasing their coverage to also cover developing countries, as I show much of the divergence in test scores happens before the points in the educational trajectories of children where they are tested by international assessments; comparable child-level panel data could substantially complement the findings of these large representative international assessments and also guard against selection on enrolment and attendance in the estimates which is likely to be an important concern in developing countries. This may also provide a more robust basis for the comparison of learning quality across countries than estimates based on the earnings of migrants to the US as problems of differential selection across countries of origin are likely to make individual country estimates unreliable.⁴¹

⁴⁰In some respects these findings are similar to structural analyses in the US but applied here to a comparative cross-country setting. See, for example, Cameron and Heckman (2001) and Keane and Wolpin (2001) who report that relaxing credit constraints at the age of 16 years does not achieve much in increasing college enrolments and that most differences in enrolment decisions seemed to predate from background factors and childhood investments.

⁴¹For example, the estimates of schooling quality used in Schoellman (2012) suggest that India has substantially better education than Vietnam. This is contrary both to test

Finally, it should be noted that the results tell us the difference in the average productivity of each completed grade in the different countries but not the sources of this differential productivity at the school level. This is an obvious area for further investigation.

References

- ANDRABI, T., J. DAS, A. I. KHWAJA, AND T. ZAJONC (2011): “Do value-added estimates add value? Accounting for learning dynamics,” *American Economic Journal: Applied Economics*, 3(3), 29–54.
- ANGRIST, J. D., P. A. PATHAK, AND C. R. WALTERS (2013): “Explaining Charter School Effectiveness,” *American Economic Journal: Applied Economics*, 5(4), 1–27.
- BEHRMAN, J. R., AND N. BIRDSALL (1983): “The quality of schooling: quantity alone is misleading,” *The American Economic Review*, 73(5), 928–946.
- BERLINSKI, S., S. GALIANI, AND P. GERTLER (2009): “The effect of pre-primary education on primary school performance,” *Journal of Public Economics*, 93(1), 219–234.
- BERLINSKI, S., S. GALIANI, AND M. MANACORDA (2008): “Giving children a better start: Preschool attendance and school-age profiles,” *Journal of Public Economics*, 92(5), 1416–1440.

scores on international assessments and to findings in this paper. Plausibly, this could reflect differential selection of migrants given that Indian migrants to the US tend to be highly skilled whereas a large number of Vietnamese refugees (who were not selected on realized human capital) were settled into the US in the aftermath of the Vietnam war.

- BLOOM, N., AND J. VAN REENEN (2007): “Measuring and explaining management practices across firms and countries,” *The Quarterly Journal of Economics*, 122(4), 1351–1408.
- (2010): “Why do management practices differ across firms and countries?,” *The Journal of Economic Perspectives*, 24(1), 203–224.
- BOND, T. N., AND K. LANG (2013): “The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results,” *Review of Economics and Statistics*, 95(5), 1468–1479.
- CAMERON, S. V., AND J. J. HECKMAN (2001): “The Dynamics of Educational Attainment for Black, Hispanic, and White Males,” *Journal of Political Economy*, 109(3), pp. 455–499.
- CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014): “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” *American Economic Review*, 104(9), 2593–2632.
- CUETO, S., AND J. LEON (2013): “Psychometric characteristics of cognitive development and achievement instruments in Round 3 of Young Lives,” *Young Lives Technical Note*, 25.
- CUETO, S., J. LEON, G. GUERRERO, AND I. MUÑOZ (2009): “Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives,” *Young Lives Technical Note*, 15.
- DAS, J., S. DERCON, J. HABYARIMANA, P. KRISHNAN, K. MURALIDHARAN, AND V. SUNDARARAMAN (2013): “School Inputs,

- Household Substitution, and Test Scores,” *American Economic Journal: Applied Economics*, 5(2), 29–57.
- DAS, J., AND T. ZAJONC (2010): “India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement,” *Journal of Development Economics*, 92(2), 175–187.
- DEMING, D. J., J. S. HASTINGS, T. J. KANE, AND D. O. STAIGER (2014): “School Choice, School Quality, and Postsecondary Attainment,” *American Economic Review*, 104(3), 991–1013.
- ENGLE, P. L., M. M. BLACK, J. R. BEHRMAN, M. CABRAL DE MELLO, P. J. GERTLER, L. KAPIRIRI, R. MARTORELL, AND M. E. YOUNG (2007): “Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world,” *The Lancet*, 369(9557), 229–242.
- ESCOBAL, J., AND E. FLORES (2008): “An assessment of the Young Lives sampling approach in Peru,” *Young Lives Technical Note*, 3.
- FIORINI, M., AND M. P. KEANE (2014): “How the Allocation of Children’s Time Affects Cognitive and Non-Cognitive Development,” *Journal of Labor Economics*, 32(4), pp. 787–836.
- FRYER, ROLAND G., J., AND S. D. LEVITT (2013): “Testing for Racial Differences in the Mental Ability of Young Children,” *American Economic Review*, 103(2), 981–1005.

- FRYER, R. G., AND S. D. LEVITT (2004): “Understanding the black-white test score gap in the first two years of school,” *Review of Economics and Statistics*, 86(2), 447–464.
- (2006): “The black-white test score gap through third grade,” *American Law and Economics Review*, 8(2), 249–281.
- FRYER, R. G., AND S. D. LEVITT (2010): “An Empirical Analysis of the Gender Gap in Mathematics,” *American Economic Journal: Applied Economics*, 2(2), 210–40.
- GLEWWE, P., E. A. HANUSHEK, S. HUMPAGE, AND R. RAVINA (2013): “School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010,” in *Education Policy in Developing Countries*, ed. by P. Glewwe. University of Chicago Press.
- GLEWWE, P., H. G. JACOBY, AND E. M. KING (2001): “Early childhood nutrition and academic achievement: a longitudinal analysis,” *Journal of Public Economics*, 81(3), 345–368.
- GLEWWE, P., AND M. KREMER (2006): “Schools, teachers, and education outcomes in developing countries,” *Handbook of the Economics of Education*, 2, 945–1017.
- GOLLIN, D., D. LAGAKOS, AND M. WAUGH (2014): “The agricultural productivity gap,” *The Quarterly Journal of Economics*, 129(2), 939–993.
- GRANTHAM-MCGREGOR, S., Y. B. CHEUNG, S. CUETO, P. GLEWWE, L. RICHTER, AND B. STRUPP (2007): “Developmental potential in the first 5 years for children in developing countries,” *The Lancet*, 369(9555), 60–70.

- GUARINO, C., M. D. RECKASE, AND J. M. WOOLDRIDGE (forthcoming):
 “Can value-added measures of teacher performance be trusted?,”
Education Finance and Policy, forthcoming.
- HANUSHEK, E. A., AND D. D. KIMKO (2000): “Schooling, Labor-Force
 Quality, and the Growth of Nations,” *The American Economic Review*,
 90(5), pp. 1184–1208.
- HANUSHEK, E. A., AND L. WOESSMANN (2008): “The role of cognitive
 skills in economic development,” *Journal of Economic Literature*, 46(3),
 607–668.
- (2012a): “Do better schools lead to more growth? Cognitive skills,
 economic outcomes, and causation,” *Journal of Economic Growth*, 17(4),
 267–321.
- (2012b): “Schooling, educational achievement, and the Latin
 American growth puzzle,” *Journal of Development Economics*, 99(2),
 497–512.
- HECKMAN, J. J., AND T. KAUTZ (2013): “Fostering and measuring skills:
 Interventions that improve character and cognition,” Working Paper
 19656, National Bureau of Economic Research.
- HECKMAN, J. J., AND S. MOSSO (2013): “The Economics of Human
 Development and Social Mobility,” *unpublished*.
- HENDRICKS, L. (2002): “How Important Is Human Capital for
 Development? Evidence from Immigrant Earnings,” *American Economic
 Review*, 92(1), 198–219.

- HSIEH, C.-T., AND P. J. KLENOW (2009): “Misallocation and manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 124(4), 1403–1448.
- KAARSEN, N. (2014): “Cross-country differences in the quality of schooling,” *Journal of Development Economics*, 107, 215–224.
- KANE, T. J., D. F. MCCAFFREY, T. MILLER, AND D. O. STAIGER (2013): “Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project.,” *Bill & Melinda Gates Foundation*.
- KANE, T. J., AND D. O. STAIGER (2008): “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation,” Working Paper 14607.
- KEANE, M. P., AND K. I. WOLPIN (2001): “The effect of parental transfers and borrowing constraints on educational attainment,” *International Economic Review*, 42(4), 1051–1103.
- KREMER, M., C. BRANNEN, AND R. GLENNERSTER (2013): “The Challenge of Education and Learning in the Developing World,” *Science*, 340(6130), 297–300.
- KUMRA, N. (2008): “An Assessment of the Young Lives sampling approach in Andhra Pradesh, India,” *Young Lives Technical Note 2*.
- MACCINI, S., AND D. YANG (2009): “Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall,” *American Economic Review*, 99(3), 1006–1026.

- McEWAN, P. J. (2013): “Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments,” *unpublished*.
- MULLIS, I. V., M. O. MARTIN, E. J. GONZALEZ, AND S. J. CHROSTOWSKI (2004): *TIMSS 2003 International Mathematics Report: Findings from IEA’s Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. ERIC.
- MUNSHI, K., AND M. R. ROSENZWEIG (2006): “Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy,” *The American Economic Review*, 96(4), 1225–1252.
- MURALIDHARAN, K. (2013): “Priorities for Primary Education Policy in India’s 12 th Five-year Plan,” in *NCAER-Brookings India Policy Forum*, vol. 2013.
- MURALIDHARAN, K., AND V. SUNDARARAMAN (2013a): “The Aggregate Effect of School Choice - Evidence from a two-stage experiment in India,” Working Paper 19440, National Bureau of Economic Research.
- (2013b): “Contract Teachers: Experimental Evidence from India,” Working Paper 19440, National Bureau of Economic Research.
- NGUYEN, N. (2008): “An assessment of the Young Lives sampling approach in Vietnam,” *Young Lives Technical Note*, 4.
- OECD (2013): *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, reading and Science (Volume I)*. OECD Publishing.

- OUTES-LEON, I., AND A. SANCHEZ (2008): “An assessment of the Young Lives sampling approach in Ethiopia,” *Young Lives Technical Note*, 1, 1–37.
- PRITCHETT, L. (2013): *The Rebirth of Education: Schooling ain’t Learning*. Brookings Institution Press for Center for Global Development, Washington, D.C.
- RIGOBON, R., AND T. M. STOKER (2007): “Estimation with Censored regressors: Basic Issues,” *International Economic Review*, 48(4), 1441–1467.
- (2009): “Bias from censored regressors,” *Journal of Business & Economic Statistics*, 27(3), 340–353.
- ROLLESTON, C. (2014): “Learning profiles and the skills gap in four developing countries: a comparative analysis of schooling and skills development,” *Oxford Review of Education*, 40(1), 132–150.
- ROLLESTON, C., Z. JAMES, AND E. AURINO (2013): “Exploring the effect of educational opportunity and inequality on learning outcomes in Ethiopia, Peru, India, and Vietnam,” *Background Paper for the UNESCO Global Monitoring Report*.
- ROTHSTEIN, J. (2010): “Teacher quality in educational production: Tracking, decay, and student achievement,” *The Quarterly Journal of Economics*, 125(1), 175–214.
- SCHOELLMAN, T. (2012): “Education Quality and Development Accounting,” *The Review of Economic Studies*, 79(1), 388–417.

- SINGH, A. (2013): “Size and Sources of the Private School Premium in Test Scores in India,” *Young Lives Working Paper*, 98.
- SINGH, A., A. PARK, AND S. DERCON (2014): “School Meals as a Safety Net: An evaluation of the Midday Meal Scheme in India,” *Economic Development and Cultural Change*, 62(2), 275–306.
- SPENCER, B. D. (1983): “On Interpreting Test Scores as Social Indicators: Statistical Considerations,” *Journal of Educational Measurement*, 20(4), pp. 317–333.
- TODD, P. E., AND K. I. WOLPIN (2003): “On the Specification and Estimation of the Production Function for Cognitive Achievement,” *The Economic Journal*, 113(485), F3–F33.
- (2007): “The production of cognitive achievement in children: Home, school, and racial test score gaps,” *Journal of Human Capital*, 1(1), 91–136.
- UN (2013): “A New Global Partnership: Eradicate poverty and Transform Economies through Sustainable Development,” The Report of the High-level Panel of Eminent Persons on the Post-2015 Development Agenda, United Nations, New York.
- VAN DER LINDEN, W. J., AND R. K. HAMBLETON (1997): “Item response theory: Brief history, common models, and extensions,” in *Handbook of Modern Item Response Theory*, ed. by W. J. Van der Linden, and R. K. Hambleton, pp. 1–28. Springer Verlag.

Appendices to Chapter 1

A. Construction of Test Scores

Introduction to Item Response Theory

Test scores used in this paper are constructed using Item Response Theory (IRT) models. IRT models, used commonly in international assessments such as PISA and TIMSS, posit a relationship between a unidimensional latent ability parameter and the probability of answering a question correctly; it is assumed that the relationship is specific to the item but is constant across individuals. Further assuming local independence, conditional on ability, between answers to different items by the same person, and across persons for the same item, it is possible to write down the likelihood function for observing the full matrix of responses, given individual-specific ability parameters and item-specific characteristics; these parameters can then be recovered based on standard maximum likelihood techniques which provide unbiased estimates of individual ability.

In this paper, following the procedure in TIMSS, I use the three-parameter logistic (3PL) model for all multiple-choice questions and the 2-PL model for items allowing open-ended responses. The 3-PL model is given by the following functional form:

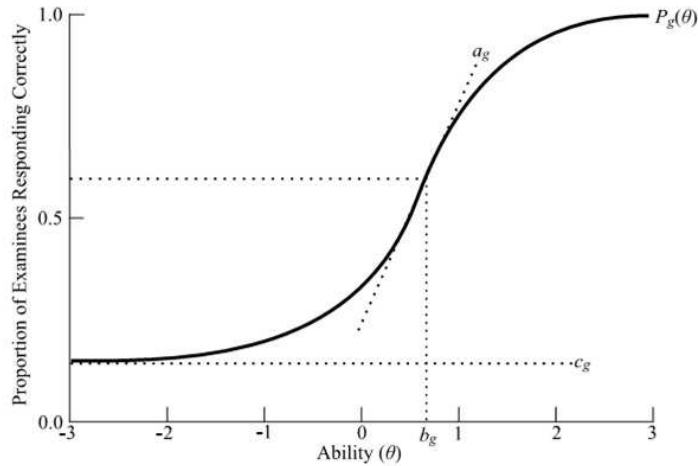
$$P_g(X_{ig} = 1|\theta_i) = c_g + \frac{1 - c_g}{1 + \exp[-1.7 \cdot a_g \cdot (\theta_i - b_g)]} \quad (10)$$

where the probability of an individual i with ability θ_i being able to correctly answer question g is given by three item-specific parameters:

the difficulty parameter b_g , the discrimination parameter a_g and the pseudo-guessing parameter c_g which accounts for the fact that with multiple choice questions even the lowest ability individual may sometimes correctly guess an answer. For the 2-PL model c_g is set to zero in which case the difficulty parameter b_g is the level of ability at which half the tested individuals would answer the question correctly.

This relationship can be depicted by plotting the relationship graphically to generate the Item Characteristic Curve, an example of which is presented in Figure A.1.

Figure A.1: Item Characteristic Curve



I used the OpenIRT suite of commands in Stata written by Tristan Zajonc to generate test scores used in this papers; specifically, I use the maximum likelihood estimates of ability for all children⁴².

⁴²Maximum Likelihood Estimates suffer from the problem that, while they provide unbiased estimates of the level of achievement, they overstate the variance. It is possible to use ‘plausible values’ estimation as used by TIMSS to generate more precise estimates of the distribution of the achievement through multiple imputation,

Testing for Differential Item Functioning

A crucial assumption underlying the use of IRT models is the absence of differential item functioning (DIF) i.e. item-specific parameters do not differ across individuals. In our application, this implies that the relationship between child ability and the probability of correctly answering a question does not differ between, say, children in Ethiopia and Vietnam. This can be a strong assumption and rules out, for example, problems due to translation of questionnaires or culture-specific framing of questions.

In order to test for the violation of the no-DIF assumption, for each item in every round of assessment, I plotted the Item Characteristic Curve based on the estimated parameters which predicts the proportion of individuals at any given ability level who will answer correctly and overlaid it with the observed proportion correct of answers at those ability levels in each country to assess if there were visible differences in Item functioning across country samples. For most items, there was no indication of DIF across countries; where any indication of DIF was visible, the item was ‘split’ in the relevant country sample i.e. treated as a separate item in the estimation of parameters and not linked to the other country samples and the IRT scores were re-estimated, following which the same procedure was repeated till no visible indications of DIF were seen. In rare cases, the probability of success did not seem to be increasing monotonically with ability (as is

as is done by TIMSS. However, these estimates are not unbiased estimates of individual ability and therefore cannot be used in the estimation of value-added models in the paper. For more details on Plausible Values methodology, please consult Mullis, Martin, Gonzalez, and Chrostowski (2004).

The brief explanation of IRT in this appendix draws upon Das and Zajonc (2010) and Van der Linden and Hambleton (1997). Readers should consult these sources for greater detail on IRT estimation.

implied by the ICC in the estimation); these items were removed from the estimation.

Table A.1 lists the items which were split following the procedures above in each of the samples and countries. Figure A.2 presents two examples of such diagnostic graphs: as is evident, the Item in Panel A does not show any evidence of DIF whereas the Item in Panel B shows distinct evidence of DIF in India⁴³.

⁴³Note that DIF in India also causes a poorer fit to the ICC in the other countries in panel B. This is noticeably improved after separating this question in India from the others in the estimation.

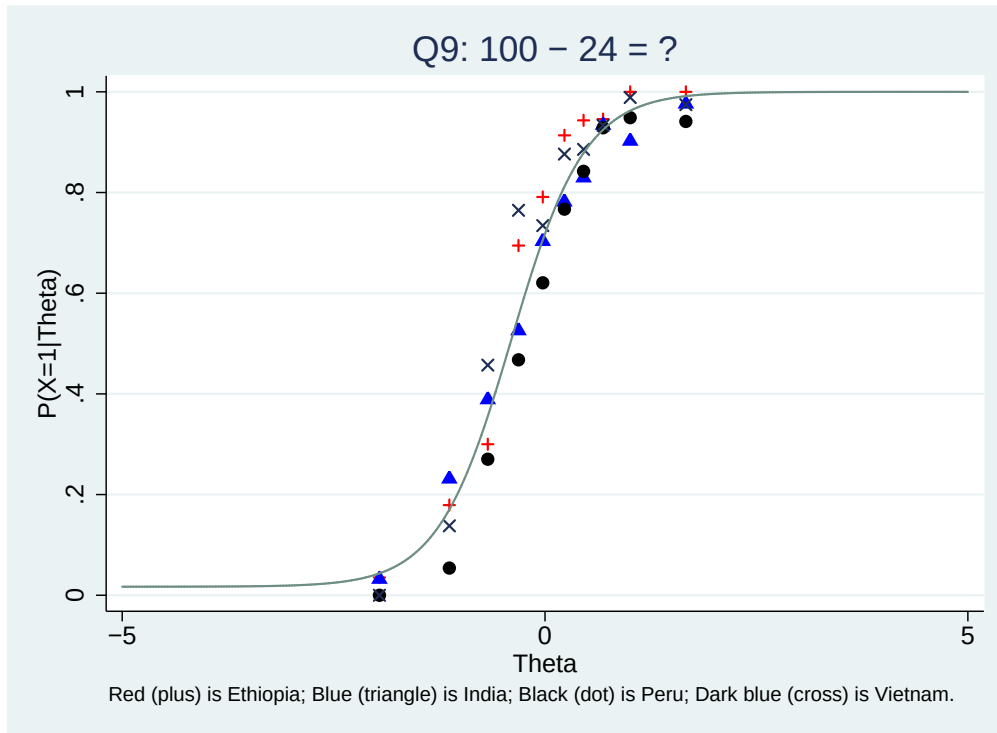
Table A.1: List of Items which were split in estimation due to DIF

Age sample	Item no.	Countries in which modified
5-years (CDA, 15 Items)	1	Ethiopia (D), India (S), Peru (D)
	3	Vietnam (S)
	6	India (S), Peru (D), Vietnam (D)
	7	Deleted in all countries
	9	Vietnam (S)
8-years (Math, 29 items)	7	Vietnam (S)
	8	Peru (S), Vietnam (S)
	9	India (S), Peru (S)
	10	Peru (S)
	15	Peru (S)
	17	Vietnam (S)
	18	Peru (S)
	20	Peru (S)
12-years (Math, 10 items)	28	India (S)
	7	Ethiopia (S), Vietnam (S)
	8	India (S)
15-years (Math, 30 items)	9	India (S), Vietnam (S)
	12	India (S)
	14	India (S)
	21	India (S), Peru (S), Vietnam (S)
	22	Ethiopia (S)
	23	Ethiopia (S), India (S)
	26	Ethiopia (S), India (S)
	27	Ethiopia (S)
	28	Ethiopia (S)

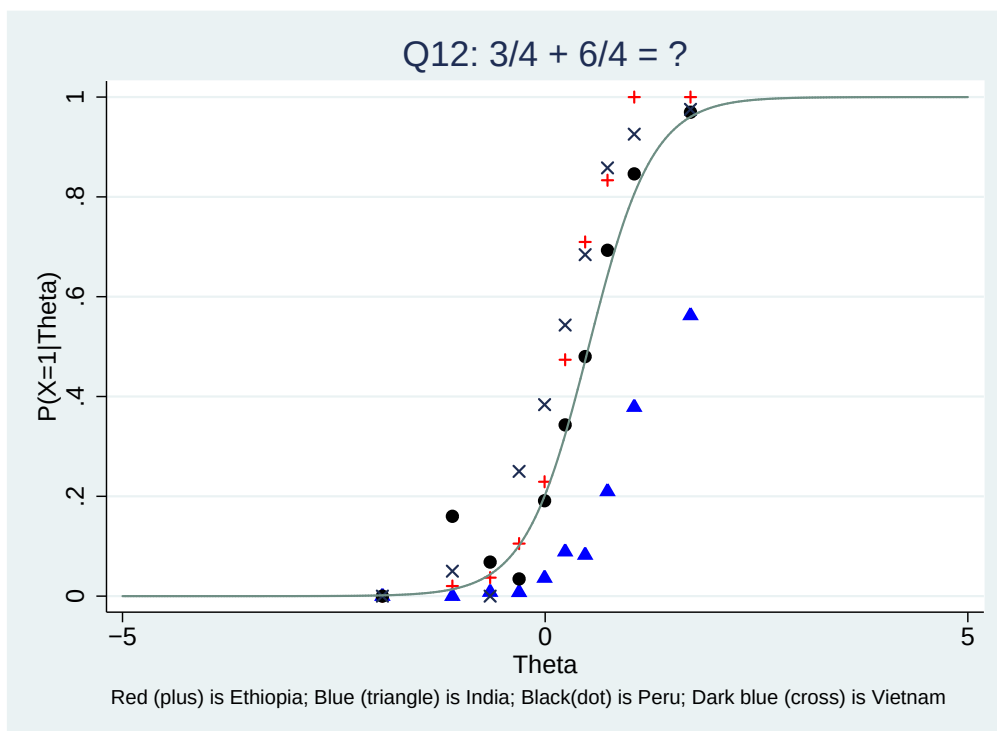
(D): Item deleted
(S): Item split in estimation

Figure A.2: Detecting Differential Item Functioning (DIF)

(a) No evidence of DIF



(b) Evidence of DIF



B. Production function estimates with flexible lag specifications

Table A.2: Allowing non-linearity in lagged achievement: 8-years old

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Ethiopia	Without time use		Vietnam	Ethiopia	With time use		Vietnam
		India	Peru			India	Peru	
Male	2.89 (5.52)	12.5*** (3.08)	8.72*** (2.28)	1.43 (2.36)	4.15 (4.84)	11.3*** (3.14)	8.89*** (2.51)	1.45 (2.58)
Eldest	4.10 (3.96)	5.19* (2.50)	8.32*** (2.81)	6.55** (2.96)	1.70 (3.69)	4.18 (2.90)	7.08** (3.01)	7.23** (3.17)
Caregiver's education level	3.65*** (0.62)	2.36*** (0.70)	2.24*** (0.50)	3.06*** (0.77)	2.70*** (0.51)	1.82*** (0.49)	2.13*** (0.48)	2.17*** (0.71)
Age in months	1.13** (0.50)	0.41 (0.47)	-0.23 (0.33)	0.15 (1.11)	1.23** (0.54)	0.52 (0.43)	-0.15 (0.31)	0.67 (0.88)
Height-for-age (2009)	8.79*** (2.59)	5.42** (2.19)	4.99** (1.84)	7.02*** (1.71)	5.02** (2.31)	4.80** (1.83)	4.62** (1.65)	4.65*** (1.50)
Wealth index (2006)	148*** (25.5)	52.3** (23.8)	14.4 (8.91)	73.8*** (20.4)	103*** (18.6)	30.0 (17.7)	15.3 (9.00)	56.4*** (18.8)
Time use (hours on a typical day)								
— doing domestic tasks					2.15 (3.34)	3.04 (4.16)	7.07*** (1.96)	-3.84 (4.06)
— doing tasks on family farm etc.					1.43 (3.44)	-13.6*** (3.52)	0.23 (1.68)	-21.3*** (5.28)
— doing paid work outside hh					-3.57 (7.88)	22.1*** (7.17)	-5.39 (3.86)	-1.85 (7.48)
— at school					11.9*** (3.33)	21.2*** (2.53)	8.93*** (2.81)	3.80 (4.09)
— studying outside school time					13.8*** (3.71)	17.7*** (4.80)	6.38*** (1.72)	2.04 (3.06)
— general leisure etc.					2.05 (3.23)	4.68* (2.49)	2.36* (1.27)	-2.47 (2.57)
— caring for others					3.30 (4.69)	1.63 (4.71)	1.80 (1.06)	-7.02 (4.80)
Highest grade completed	40.6*** (4.74)	27.1*** (2.08)	33.4*** (3.72)	60.0*** (14.5)	28.3*** (4.56)	25.1*** (1.64)	32.5*** (3.65)	54.8*** (10.9)
Lagged test score	-0.89** (0.40)	-0.43 (0.27)	0.32 (0.43)	-1.08** (0.39)	-0.54 (0.34)	-0.43* (0.24)	0.31 (0.39)	-1.08** (0.39)
Lagged score, squared	0.0020** (0.00081)	0.0013** (0.00048)	-0.000041 (0.00071)	0.0024*** (0.00070)	0.0012* (0.00070)	0.0012*** (0.00043)	-0.000022 (0.00064)	0.0022*** (0.00075)
Lagged score, cubed	-1.23e-06** (5.19e-07)	-9.16e-07*** (2.77e-07)	-1.80e-07 (3.87e-07)	-1.54e-06*** (4.14e-07)	-8.02e-07* (4.48e-07)	-8.49e-07*** (2.60e-07)	-1.87e-07 (3.53e-07)	-1.36e-06*** (4.59e-07)
Constant	349*** (58.7)	385*** (69.7)	342*** (78.7)	532*** (97.7)	220*** (63.6)	177** (67.1)	257*** (74.2)	520*** (100)
Observations	1,835	1,892	1,888	1,907	1,834	1,892	1,881	1,858
R-squared	0.343	0.280	0.355	0.441	0.411	0.368	0.381	0.461

Robust standard errors in parentheses. Standard errors are clustered at site level. *** p<0.01, ** p<0.05, * p<0.1
Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age.
Coefficients should be compared to Table 10 which is the analogous specification entering lagged achievement linearly

Table A.3: Allowing non-linearity in lagged achievement: 15-years old

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Ethiopia	Without time use		Vietnam	Ethiopia	With time use		Vietnam
		India	Peru			India	Peru	
Male	24.6*** (7.06)	21.9*** (3.24)	-4.39 (3.35)	-6.01 (4.24)	26.4*** (6.35)	18.8*** (4.67)	-2.16 (3.52)	-3.08 (4.00)
Eldest	-1.54 (7.59)	2.82 (3.72)	2.70 (2.93)	2.62 (3.49)	-2.47 (7.20)	3.60 (4.01)	3.55 (3.08)	0.61 (3.60)
Caregiver's education level	-1.17* (0.65)	2.00*** (0.49)	0.70** (0.28)	2.17** (0.87)	-1.17* (0.66)	1.16** (0.54)	0.59** (0.26)	1.72** (0.82)
Age in months	-0.33 (0.90)	-0.35 (0.46)	-0.65* (0.36)	-0.93 (0.79)	-0.19 (0.89)	0.18 (0.46)	-0.41 (0.29)	-0.29 (0.79)
Height-for-age (2009)	1.60 (2.89)	-0.027 (2.64)	1.69 (1.68)	2.39 (2.63)	1.60 (2.78)	0.78 (2.54)	2.18 (1.65)	2.63 (2.81)
Wealth index (2006)	67.9** (23.9)	35.2** (15.4)	-5.67 (7.92)	63.0** (23.5)	55.4** (22.8)	27.5* (15.3)	-8.87 (7.77)	53.1** (23.1)
Time use (hours on a typical day)								
— doing domestic tasks					4.56 (3.18)	3.08 (2.35)	2.00 (1.32)	1.17 (3.52)
— doing tasks on family farm etc.					3.99 (3.40)	-1.00 (1.98)	1.25 (1.51)	0.0012 (2.32)
— doing paid work outside hh					4.11 (3.37)	-1.52 (1.65)	1.44 (1.81)	-0.19 (1.63)
— at school					8.51** (3.17)	3.60* (1.84)	4.64*** (1.33)	4.58 (2.81)
— studying outside school time					10.7*** (2.67)	5.73*** (1.95)	3.48*** (1.03)	1.02 (2.03)
— general leisure etc.					6.00 (3.61)	0.14 (1.46)	1.54 (1.20)	-1.20 (2.15)
— caring for others					7.99** (3.71)	-8.84** (4.02)	0.0050 (1.13)	-2.58 (3.71)
Highest grade completed	18.8*** (2.23)	12.4*** (1.98)	11.4*** (2.65)	9.75*** (2.16)	16.0*** (2.44)	7.41*** (1.36)	8.83*** (2.50)	8.22*** (2.25)
Lagged test score	-6.96*** (1.84)	-3.27* (1.76)	0.90 (2.06)	-2.14 (3.77)	-7.11*** (1.97)	-2.65 (1.60)	0.60 (2.09)	-0.71 (3.49)
Lagged score, squared	0.017*** (0.0038)	0.010*** (0.0036)	0.00099 (0.0039)	0.0068 (0.0075)	0.017*** (0.0040)	0.0089** (0.0033)	0.0016 (0.0039)	0.0035 (0.0069)
Lagged score, cubed	-0.000012*** (2.49e-06)	-8.48e-06*** (2.40e-06)	-1.90e-06 (2.44e-06)	-5.37e-06 (4.88e-06)	-0.000012*** (2.66e-06)	-7.30e-06*** (2.21e-06)	-2.28e-06 (2.44e-06)	-3.06e-06 (4.45e-06)
Constant	1,215*** (343)	517* (287)	118 (356)	655 (631)	1,120*** (340)	373 (277)	99.5 (361)	358 (568)
Observations	964	970	656	964	963	969	656	925
R-squared	0.445	0.563	0.547	0.437	0.462	0.599	0.575	0.456

Robust standard errors in parentheses. Standard errors are clustered at site level. *** p<0.01, ** p<0.05, * p<0.1

Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age.

Coefficients should be compared to Table 11 which is the analogous specification entering lagged achievement linearly

C. Robustness of estimates at 15 to censoring of achievement at age 12

Table A.4: Production function estimates at 15-years using Bayesian EAP lagged scores

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Without time use				With time use			
	Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Male	24.4*** (7.02)	20.9*** (4.27)	-3.32 (3.69)	-5.43 (4.30)	26.1*** (6.24)	18.4*** (5.77)	-1.13 (3.74)	-2.68 (3.93)
Eldest	-1.48 (7.68)	2.25 (3.49)	2.70 (2.68)	2.38 (3.38)	-2.44 (7.32)	3.25 (3.70)	3.37 (2.95)	0.17 (3.46)
Caregiver's education level	-1.34* (0.67)	1.70*** (0.50)	0.51 (0.35)	2.03** (0.88)	-1.34* (0.69)	0.86 (0.51)	0.40 (0.31)	1.62* (0.83)
Age in months	-0.39 (0.90)	-0.98* (0.49)	-0.80* (0.42)	-1.07 (0.81)	-0.26 (0.88)	-0.36 (0.51)	-0.56 (0.33)	-0.45 (0.81)
Height-for-age (2009)	1.62 (2.87)	-0.50 (2.55)	1.38 (2.09)	2.39 (2.59)	1.61 (2.76)	0.49 (2.48)	1.72 (1.99)	2.61 (2.76)
Wealth index (2006)	69.1*** (23.7)	32.6* (17.7)	-3.92 (8.11)	62.9** (23.5)	57.0** (22.8)	24.9 (16.5)	-6.05 (8.10)	53.5** (23.1)
<i>Time use (hours on a typical day)</i>								
— doing domestic tasks					4.25 (3.14)	3.08 (2.76)	2.37 (1.43)	1.05 (3.69)
— doing tasks on family farm etc.					3.71 (3.32)	-2.02 (2.20)	0.66 (1.48)	-0.35 (2.44)
— doing paid work outside hh					3.83 (3.36)	-3.20 (1.90)	2.20 (1.86)	-0.17 (1.66)
— at school					8.10** (3.10)	3.15 (1.99)	5.30*** (1.57)	4.32 (2.87)
— studying outside school time					10.5*** (2.72)	4.72** (1.88)	2.66** (1.09)	0.65 (2.12)
— general leisure etc.					5.62 (3.57)	-0.23 (1.61)	1.52 (1.15)	-1.42 (2.15)
— caring for others					7.31* (3.72)	-9.60** (4.11)	0.85 (1.05)	-2.64 (3.80)
Lagged math scores (2006)	0.34*** (0.035)	0.43*** (0.047)	0.22*** (0.023)	0.33*** (0.035)	0.34*** (0.034)	0.39*** (0.040)	0.22*** (0.021)	0.29*** (0.035)
Highest grade completed	19.6*** (2.06)	16.8*** (1.68)	15.3*** (3.32)	10.5*** (2.26)	16.9*** (2.31)	10.9*** (1.24)	12.6*** (2.95)	9.01*** (2.45)
Constant	226 (167)	274*** (91.8)	444*** (62.9)	442*** (130)	120 (167)	211* (103)	377*** (45.3)	357** (131)
Observations	964	970	656	964	963	969	656	925
R-squared	0.441	0.532	0.494	0.432	0.457	0.574	0.525	0.451

Robust standard errors in parentheses. Standard errors are clustered at site level. *** p<0.01, ** p<0.05, * p<0.1

Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age. The dependent variable is the maximum likelihood IRT estimate as in all previous tables.

Lagged achievement (at the age of 12) is the Bayesian EAP test score which is more robust to ceiling and floor effects.

Coefficients should be compared to Table 11 which is identical but for using MLE scores as the lagged achievement measure.

D. Correcting for measurement error in lagged achievement

Table A.5: Estimates correcting for measurement error: 8-years old

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var: Mathematics score (2009)							
	Without time use				With time use			
	Ethiopia	India	Peru	Vietnam	Ethiopia	India	Peru	Vietnam
Highest grade completed	40.9*** (4.67)	27.4*** (2.03)	33.6*** (3.60)	60.9*** (14.6)	28.4*** (4.48)	25.4*** (1.62)	32.6*** (3.55)	55.2*** (10.9)
Male	1.18 (4.84)	14.1*** (2.67)	8.97*** (2.33)	1.65 (2.54)	3.89 (3.91)	13.4*** (2.83)	9.43*** (2.46)	1.63 (2.51)
Eldest	-1.57 (3.75)	2.36 (2.87)	7.19*** (2.35)	3.57 (2.58)	-2.17 (3.99)	2.13 (2.95)	5.88** (2.48)	4.89* (2.79)
Caregiver's education level	2.55*** (0.49)	1.73*** (0.50)	1.58*** (0.32)	2.57*** (0.57)	1.98*** (0.44)	1.47*** (0.46)	1.51*** (0.31)	2.17*** (0.64)
Age in months	0.59 (0.39)	-0.073 (0.32)	-0.43 (0.32)	-0.79 (1.10)	0.55 (0.44)	0.010 (0.31)	-0.43 (0.32)	-0.31 (0.86)
Height-for-age (2009)	6.72*** (2.06)	3.41* (1.99)	3.05** (1.45)	4.50*** (1.45)	3.28* (1.93)	3.59** (1.76)	2.76** (1.36)	3.24** (1.35)
Wealth index (2006)	60.8*** (14.2)	68.0*** (16.9)	18.4*** (5.22)	40.3*** (14.8)	41.9*** (12.2)	48.0*** (14.6)	18.4*** (5.47)	31.7** (13.5)
Time use (hours on a typical day)								
— doing domestic tasks					2.60 (2.61)	5.17 (3.74)	4.50*** (1.43)	-1.49 (3.61)
— doing tasks on family farm etc.					0.69 (2.83)	-18.3*** (4.05)	1.76 (2.31)	-12.0** (4.68)
— doing paid work outside hh					-6.12 (7.28)	17.1*** (6.33)	3.25 (5.36)	5.11 (11.2)
— at school					10.8*** (3.44)	16.5*** (2.82)	6.62*** (2.35)	8.73** (4.42)
— studying outside school time					12.7*** (2.81)	11.5*** (2.53)	6.07*** (1.15)	8.85*** (2.29)
— general leisure etc.					1.01 (2.46)	2.01 (1.44)	1.02 (1.01)	1.61 (1.67)
— caring for others					1.90 (3.63)	0.13 (4.18)	2.10** (0.89)	-1.92 (3.16)
Lagged CDA scores (2006)	0.15** (0.069)	0.32*** (0.045)	0.27*** (0.027)	0.20*** (0.065)	0.11* (0.057)	0.27*** (0.044)	0.26*** (0.030)	0.15** (0.064)
Constant	299*** (41.0)	217*** (33.5)	339*** (28.4)	426*** (83.0)	228*** (54.2)	94.7** (42.1)	281*** (36.8)	345*** (65.2)
Observations	1,821	1,821	1,848	1,708	1,820	1,821	1,842	1,662
R-squared	0.409	0.343	0.323	0.486	0.466	0.413	0.346	0.507
Kleibergen-Paap F-statistic	72.3	81.2	201	56.2	73.6	82.7	193	53.0

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses. Standard errors are clustered at site level. Test scores are IRT scores normalized to have a mean of 500 and SD of 100 in the pooled four-country sample at each age. Lagged CDA scores are instrumented using scores on the adapted Peabody Picture Vocabulary test in 2006 to correct for measurement error. Coefficients should be compared to Table 10.

Chapter 2

Private school effects in urban and rural India: Panel estimates at primary and secondary school ages

Abstract

I present the first value-added models of learning production in private and government schools in India using unique panel data from Andhra Pradesh state. I examine the heterogeneity in private school value-added across different subjects, urban and rural areas, medium of instruction, and across age groups. Further, I also estimate private school effects on children's self-efficacy and agency. In rural areas, I find a substantial positive effect (>0.5 SD) of private schools on English, no effect on mathematics and heterogeneous effects on Telugu for 8–10 year old students; at 15 years, there are modest effects (<0.2 SD) on mathematics and Telugu receptive vocabulary. I find no evidence of a positive effect in urban areas or on psychosocial skills. Results on comparable test domains and age groups correspond closely with, and further extend, estimates from a parallel experimental evaluation.

1 Introduction

The share of private schools in total enrolment has risen substantially across both urban and rural areas in India in the past 15 years (Kingdon, 2007), now accounting for about 30% of enrolment in rural areas (Pratham, 2013); students in these schools perform much better on average in test scores (Muralidharan and Kremer, 2008); and frequently it seems that private schools achieve this better performance even with much lower expenditure per pupil than government schools (Desai, Dubey, Vanneman, and Banerji, 2008).

In this paper, I answer three questions using a unique longitudinal dataset collected by the Young Lives study in Andhra Pradesh state, which has tracked two cohorts of children through multiple rounds of data collection at household and school level between 2002 and 2011. First, I estimate value-added models (VAMs) of learning achievement to evaluate whether, and to what extent, the better test performance of children in private schools is attributable to schools. Second, I examine whether the impact of private schools on test scores is heterogeneous across different tests, different age groups, different languages of instruction and across urban and rural areas. Third, I attempt to assess if private schools have a differential impact on psychosocial (‘non-cognitive’) skills of children, specifically their self-efficacy and agency.

These questions are central to understanding the implications of the rise of private schools for the educational sector in India. Establishing whether the private school premium is causal, and whether it is heterogeneous, is essential for understanding whether the rapid proliferation of private

schools can be expected to improve abysmally low learning levels in Indian schools and, if so, in which domains and by how much; this is also essential for understanding the likely implications for inequality in test scores in India, which is already among the highest in the world (Das and Zajonc, 2010). Finally, an assessment of the impact of private schools on psychosocial skills is relevant in light of a large interdisciplinary literature, including recent studies in economics, that documents their importance both in the production of learning skills and on later life outcomes (e.g. Bandura, 1982, 1993; Cunha and Heckman, 2008; Cunha, Heckman, and Schennach, 2010).

In rural areas, where private schools account for about a quarter of the total enrolment in our data, I find that value-added in test scores for students in private schools is substantially greater in English compared to government school students, moderately greater in receptive vocabulary and no worse in Mathematics between the ages of 8 and 10 years; in Telugu, the local language, children in English-medium private schools do worse than government school students but children in Telugu-medium private schools perform as well while still out-performing government school students in English. At the age of 15 years, children in private schools significantly outperform government school children in Mathematics and Telugu receptive vocabulary, although estimated effect sizes remain relatively modest and only about 20 to 40 per cent of the (unconditional) cross-sectional difference in test scores. In urban areas, I find no evidence of a causal private school effect on test scores.

Private school students display higher cross-sectional levels of agency and self-efficacy than government school students. However, these differences

invariably disappear upon inclusion of socio-economic controls and lagged achievement; in short, I do not find any evidence of a causal private school effect on psychosocial skills.

This paper makes significant contributions to the literature on the impact of private schools in India. It is the first paper in this literature to use individual-level panel data to estimate value-added models of achievement; it is the only study to cover both primary and secondary school aged children across urban and rural areas; in comparison to previously published work (discussed in Section 2), the analysis makes use of considerably more detailed test measures and background information in each survey wave; and finally, it is the first study to look at differences in the psychosocial skills across private and government schools.

I attempt in this paper to also add to a recent and growing literature, mostly from studies in the United States, on the robustness and reliability of value-added approaches to modelling achievement production. In particular, the richness of the data allow me to test some of the assumptions implicit in value-added models and quantify the extent to which any violations may bias estimated private school effects in these data. Using the Raven's test from two periods ago as a proxy, I show that bias due to selection on ability does not appear to significant. I further show that controlling for lagged parental assessments of a child's academic performance and for parental aspirations with regard to a child's education does not alter the size or significance of estimated private school effects from value-added models. This indicates that dynamic sorting, which may be caused due to decision-makers having access to different or more detailed

information about children than just their previous test scores, also does not seem to be a major source of bias in these estimates.

Furthermore, part of the data used in this paper are contemporaneous to data collected by Muralidharan and Sundararaman (2013a, MS hereafter) for children of the same age, in the same state and tested on partly the same domains of learning. The MS study offers experimental variation, induced through the randomized assignment of school vouchers in a representative set of communities in Andhra Pradesh, and thus offers an ideal comparison for results in this paper to assess the presence and extent of systematic bias that may still be present in value-added models. As I document, the pattern of causal effects reported here, on a comparable set of indicators for a comparable cohort of children, is very similar to the MS study; this is, to my knowledge, the first comparison of experimental and value-added estimates for the effect of selection into different schools in a developing country and the first such comparison using independently drawn samples in any setting. In addition to documenting similar results for one sample of children, I extend the results of MS substantially by presenting results on older children of secondary school age, on children in urban areas and on non-curricular test domains (receptive vocabulary) and psychosocial skills. The MS study is focused exclusively on the curricular outcomes of children in rural areas aged about 8 to 10 years at the time of testing.

These results have important implications. Combined with the significantly lower per-pupil expenditure in private schools, discussed in Section 2, this indicates that private schools are considerably more productive than government schools on average. The much better performance of private school students in English may plausibly contribute a significant labour

market premium for these children in the future; recent evidence using nationally representative data from 2005 suggests an increase in hourly wages by 13 per cent for men who can speak a little English and up to 34 per cent for those who can speak it fluently (Azam, Chin, and Prakash, 2013). However, the insignificance or relatively modest size of the private school premium in most dimensions (with the exception of English) indicates that the spread of private schooling alone, without concomitant reforms across the education sector, will not lead to very appreciable improvements in the low levels of learning in Indian schools as measured by achievement in mathematics or the ability to read and write.

The robust performance of value-added models, in a developing country setting with decidedly non-random selection across school types, may have methodological and practical relevance. While experimental evidence, preferably on representative samples and with minimal attrition, remains very desirable for measuring the causal effects of different educational inputs and interventions, such data are unlikely to be always available or always feasible to collect; results in this paper support the reliability of value-added estimates using panel data in these settings.

Results in this paper may resonate more broadly than merely Andhra Pradesh or the Indian context. Low-fee private schools have increased their share in enrolment across several developing countries; in many countries in Latin America, Asia and Africa, they also seem to outperform government schools in test scores (Jimenez, Lockheed, and Paqueo, 1991; Alderman, Orazem, and Paterno, 2001; Andrabi, Das, Khwaja, and Zajonc, 2011; Bold, Kimenyi, Mwabu, and Sandefur, 2011). Given the similarity

in findings on several dimensions of schooling across South Asia and these different contexts, evidence from India may have relevance for these contexts as well.¹ Similarly, the methodological question about the presence and extent of any bias in value-added models is also relevant across different contexts.

The rest of this paper is structured as follows: Section 2 presents the background and context of the schooling sector in India; Section 3 presents the data; Sections 4 and 5 present the empirical specifications and results from VAMs exploring whether the private school effect is causal and whether it is heterogeneous; Section 6 discusses additional concerns and robustness; Section 7 explores the impact of private schools on psychosocial skills; Section 8 concludes.

2 Private and Government schools in India

As noted previously, the share of the private sector in total enrolment especially at the primary level has expanded very rapidly and a large

¹For instance, the insignificant or modest gains from input-based measures seems to hold across South Asian and African contexts (see Kremer, Brannen, and Glennerster, 2013; McEwan, 2013; Glewwe, Hanushek, Humpage, and Ravina, 2013) as do institutional issues of teacher absenteeism (Chaudhury, Hammer, Kremer, Muralidharan, and Rogers, 2006), low rates of attainment (Glewwe and Kremer, 2006) and equal or greater productivity of contract or para teachers (Atherton and Kingdon, 2010; Duflo, Dupas, and Kremer, 2012; Muralidharan and Sundararaman, 2013b). For example, comparing their results on an extra contract teacher in Kenya to those of Muralidharan and Sundararaman (2013b) from the state of Andhra Pradesh, Duflo, Dupas and Kremer (2012) note that “[Muralidharan and Sundararaman] find average test score gains very similar to those we observe, as well as reductions in civil-service teacher presence, suggesting that very similar forces may be at play in a different geographic and institutional context.” (p. 24)

While this study cannot itself claim external validity beyond the state of Andhra Pradesh in India, it is plausible that the results have some relevance for other Indian states and perhaps more broadly.

literature finds significant difference in the test scores of children in these schools when compared to government school students. It has also been shown that government school teachers, although better-paid and more qualified than private school teachers, are also much more likely to be absent and much less likely to be teaching when in school (see e.g. Desai, Dubey, Vanneman, and Banerji 2008; Kingdon 2007; Muralidharan and Kremer 2008).

Several studies have noted that per-pupil spending is considerably higher in government schools compared to private schools. Kingdon (1996) reports, using data on schools in Lucknow from 1990, that per-pupil expenditure in government schools was roughly double the per-pupil expenditure in private unaided schools. This ratio has since multiplied; Muralidharan and Sundararaman (2013a), using data from 2008, report annual cost per-child of Rs. 8390 in government schools and Rs. 2324 in private schools in Andhra Pradesh. This difference in per-child spending is largely reflects much higher salaries for permanent civil service (government) teachers compared to private school teachers and has increased due to the much higher growth in government salaries. Kingdon (2010) discusses this issue at considerable length with an overview of evidence from the last two decades. She also notes that such differences must have increased considerably since the Sixth Pay Commission recommendations have near-doubled the average salaries of teachers in government schools.

Assessing the causal contribution of private schools to the learning outcomes of pupils is beset with serious problems of endogenous selection: there are systematic differences in observable characteristics of students in the two sectors and it is plausible that these may extend also to

unobservable characteristics. Studies in the Indian context have adopted a series of econometric techniques to correct for this source of bias.² The results in most cases seem to indicate that there is, in fact, a ‘private school premium’ in test scores which persists even when issues of selection have been dealt with as far as possible.

However, most studies previously available in the literature use only cross-sectional variation in test scores and a limited range of characteristics of children, schools and households to arrive at their estimates of the private school premium. An exception is the analysis presented by French and Kingdon (2010) who create a village-level panel by looking at villages that were included in multiple rounds of Pratham’s ASER survey between 2005 and 2007; this enables them to use village and year fixed effects. To my knowledge, with the exception of the contemporaneous MS study, no other study has been able to utilize individual-level panel variation in the data. The identification of private school effects in previous work in this literature, while perhaps the best that can be achieved given the data, remains vulnerable to several sources of endogeneity.³

²Specifically, these studies have used the following approaches: controlling for observed background characteristics of children (Muralidharan and Kremer, 2008; Kingdon, 1996; French and Kingdon, 2010; Desai, Dubey, Vanneman, and Banerji, 2008); running models with village fixed effects to isolate village-level confounders (French and Kingdon, 2010); with household fixed effects in cross-sectional data (French and Kingdon, 2010); with propensity score matching (Chudgar and Quin, 2012); and finally, through the use of Heckman selection models (Kingdon, 1996; Desai, Dubey, Vanneman, and Banerji, 2008).

³Ordinary Least Squares (OLS) regressions and propensity score matching estimators conditioning for limited background characteristics cross-sectionally are unlikely to have observed all relevant dimensions in which these children differ; within-village comparisons neglect the potential bias caused due to unobserved characteristics that lead households in the same village to making different choices regarding the enrolment of their children; household fixed effects remain vulnerable to differential enrolment within the household being related to either unobserved ability differences across children or, even more plausibly, to other unobserved differences in complementary investments; finally, variables used to control for selection in these studies using Heckman selection-correction estimators are unlikely to satisfy necessary exclusion restrictions.

The most convincing results on the effect of private schools in India are those from the MS study from the state of Andhra Pradesh, which is also the state in which the data used in this paper are collected. The MS study offered school vouchers through random assignment to children in the last year of pre-school (kindergarten) and Grade 1 for the entire duration of primary schooling up till Grade 5 which could be used to attend any private school in the village; this allows for clean identification of the magnitude of any private school effect. Results from the study indicate that children in private schools perform better in English and Hindi and no worse in Mathematics and Telugu even though up to 40% less instruction time is dedicated to these subjects in private schools than in government schools. The Young Lives data do not have Hindi test scores; however, for Mathematics, Telugu and English tests administered to a comparably aged cohort I document similar estimates of the private school effect.

3 Data

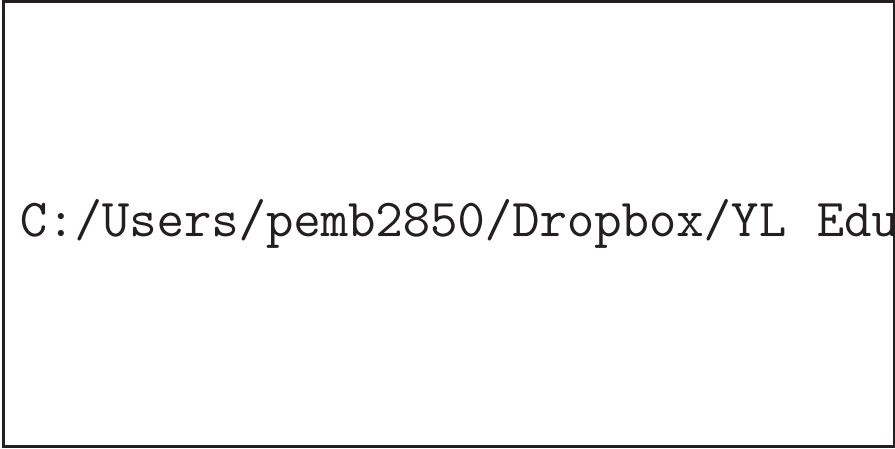
3.1 Sampling

The data I use in this study were collected by the Young Lives study between 2002 and 2011 in the state of Andhra Pradesh.⁴ Andhra Pradesh is the fourth-largest state in India by area and had a population of

For example, Desai, Dubey, Vanneman, and Banerji (2008) use the presence of a private school in a village as a factor predicting selection into private schools but not test scores; this exclusion restriction is almost certainly untenable as villages that have a private school will differ from villages that do not. In fact such a pattern has clearly been documented by Pal (2010) using the PROBE dataset covering five Indian states.

⁴Young Lives is a longitudinal study of child poverty which follows two cohorts of children in four countries: Ethiopia, Andhra Pradesh state (India), Peru and Vietnam. For details, please visit www.younglives.org.uk

Figure 1: Timing of interviews in Young Lives



C:/Users/pemb2850/Dropbox/YL Education/Output/D

over 84 million in 2011. It is divided into three regions – Coastal Andhra, Rayalaseema and Telangana – with distinct regional patterns in environment, soil and livelihood patterns. Administratively the state is divided into districts, which are further sub-divided into sub-districts (mandals) which are the primary sampling units within our sample.⁵

The Young Lives study in Andhra Pradesh has collected data on two cohorts of children: 1008 children born between January 1994 and June 1995, and 2011 children born between January 2001 and June 2002. Data was collected from children and their families using household visits in 2002, 2007 and 2009/10. The study also collected extensive data through visits to the schools of a randomly selected sub-sample of the younger cohort in 2011. Figure 1 presents graphically the timings of data collection and the

⁵The Young Lives sample is distributed across the three main regions and covers about 100 communities (villages or urban wards) across 20 sub-districts (mandals). A careful comparison with representative data for Andhra Pradesh shows that the data in the Young Lives sample contains similar variation across comparable measures: a detailed explanation of the sampling methodology and the comparison of the characteristics of the Young Lives sample with the DHS sample on a range of observed characteristics is reported in Kumra (2008). In 2014, the state of Andhra Pradesh was bifurcated with the creation of the new state of Telangana; throughout this paper, Andhra Pradesh refers to the undivided state.

age of the children at the time of the data collection.⁶ Attrition rates in the data have been kept very low – 1932 children (96 per cent) in the younger cohort and 976 children (97 per cent) in the older cohort are still in the sample in 2009.⁷ This has been achieved in part by following children whose households migrated from their original communities to their destination of migration.

3.2 Data collected through household visits

Extensive test data were collected from children in the sample in all rounds of the survey. The tests differed in their focus on which dimension of cognitive achievement they attempted to capture and how closely they related to the formal school curriculum in Andhra Pradesh; often, different tests were administered to children across rounds in order to ensure that they were appropriate for the age and the stage of education that the children were in. Box 1 lists the different test measures that are used in this paper.⁸

⁶The interviews were usually carried out over a period of four to six months for the bulk of the sample. The timing of interviews given in Figure 1 corresponds to the end-period for the majority of the interviews which did not involve tracking children to different communities.

⁷Of the children who left the sample between 2002 and 2009, 36 children in the younger cohort and 5 in the older cohort had died between the rounds; 4 children in the younger cohort and 10 in the older cohort left as they did not want to be part of the survey anymore; the whereabouts of the rest could not be tracked. Most of this attrition occurred between the 2002 and 2006 rounds. Details of the attrition including characteristics of the individuals lost to attrition and implications for the representativeness of the sample are addressed at length in ?.

⁸For precise details of the contents of the tests, as well as the validation for use in Andhra Pradesh, please consult Cueto, Leon, Guerrero, and Muñoz (2009), Cueto and Leon (2013) and the Young Lives questionnaires which are available at www.younglives.org.uk.

Box 1. Cognitive Tests in Young Lives

COHORT	ROUND 1 (2002)	ROUND 2 (2007)	ROUND 3 (2010)	SCHOOL SURVEY (2011)
Older Cohort	8 years old	12 years old	15 years old	
	Raven's test	PPVT Mathematics	PPVT Mathematics Telugu (cloze)	[Not covered]
Younger Cohort	6–24 months old	5 years old	8 years old	9 years old
		PPVT CDA Quantitative	PPVT Writing Mathematics	Mathematics Telugu English

PPVT refers to the Peabody Picture Vocabulary Test - III adapted for use in Telugu.

CDA refers to the Cognitive Development Assessment quantitative sub-scale.

Scores on all tests used in this paper, with the exception of the Raven's test, were generated using Item Response Theory (IRT) models. The use of IRT models is standard in education assessments and presents significant advantages: it allows for the accounting of difficulty of different items and, where the same test (or a sub-set) was administered over time, it allows for the computation of scores from the repeated tests on the same scale.⁹

Tests in which the same items were administered were calibrated together

⁹For a detailed explanation of IRT models, please consult Das and Zajonc (2010) and Van der Linden and Hambleton (1997). I used the OpenIRT suite of commands in Stata written by Tristan Zajonc to generate the maximum likelihood scores used in this papers. A core assumption of IRT models is that item parameters (e.g. difficulty) do not differ across sub-groups. This is not an assumption that is maintainable across different languages as difficulty levels may plausibly have changed during translation and therefore, in the case of the PPVT and the cloze test ('fill-in-the-blanks') administered to the older cohort in 2009/10, I am constrained in only using the test scores of children who took the tests in Telugu; these account for over 90 per cent of the sample in each of the tests.

which allows them to be put on the same scale. Test scores are normalized to have a mean of 0 and a standard deviation of 1.¹⁰

The tests used in Young Lives are much more comprehensive in the domains of learning they capture and offer more variation than tests in most previous studies in the literature, which is a considerable strength of the data. In particular, the ASER data collected by Pratham and the India Human Development Survey 2005 (used in French and Kingdon 2010, Desai et. al. 2008 and Chudgar and Quin 2012) have only very basic test measures, on a more limited set of domains, which are not best-suited to capturing variation across the ability spectrum.

Data collection in 2002, 2007 and 2010 was at the households of the children. This data has particularly rich information about the socio-economic background of the children's households, parental expectations/aspirations for the children, and also detailed child-specific data. In the interest of clarity, I will explain individual variables being used in the estimation in the different empirical sections at the point at which they are being employed.

¹⁰The same items were administered in the PPVT in both rounds and a subset of items from the Round 3 math tests for the younger cohort were repeated in the school survey. In the case of tests in different rounds which were calibrated together, I have normalized scores by cohort to have a mean of zero in the first period in which the test is administered. Math scores for the older cohort in Rounds 2 and 3 cannot be linked to a common scale due to the unavailability of adequate link items administered in both rounds.

3.3 Data collected from schools

In 2011, the Young Lives study visited a random sub-set of schools being attended by children in the younger cohort.¹¹ The sampling frame consisted of all the Younger Cohort (YC) children who were still enrolled in school in Round 3 (2009) and were going to school within Andhra Pradesh.¹² The final sample includes 952 children across 247 schools.¹³

The school-level survey was conducted between December 2010 and March 2011, i.e. in the school year immediately after the third wave of household-level data collection. The survey administered questionnaires to all Young Lives sample children in the school which included a battery of attitudinal items (used to generate measures of psychosocial skills). Sample children were administered tests in mathematics, Telugu and English.

¹¹It was not possible to visit all schools due to budgetary and logistical constraints. In total, 807 different schools were being attended by children in this cohort in 2009, 538 of those attended by only one Young Lives sample child; logistical and funding considerations limited the size of the sample targeted to 250 schools.

¹²Sample children outside Andhra Pradesh were excluded from the frame as tracking them was unfeasible and because all questionnaires, tests and procedures were designed keeping the state education system in context; this left 1880 children in the sampling frame.

¹³The sampling was carried out within strata defined on whether the school was in an urban or a rural area, whether it was private or public and whether it was recognized or unrecognized, yielding a total of six strata. In each stratum, a pre-determined number of children were drawn randomly and all other Young Lives YC children in the school were covered as well. Where the child(ren) enrolled in a particular school had shifted schools since 2009, they were dropped from the school-based survey and were not followed to their new school unless the new school was also already in the sample. See ? for a full description of the sampling procedure. Descriptive statistics and regressions for this sample are unweighted in this paper as appropriate survey weights are not archived in the Young Lives public use datasets.

4 Empirical Framework

Analysis in this paper, based on observational longitudinal data, will rely on value-added models of achievement production. These models have been widely used in the literature on determinants of learning with several recent papers discussing their conceptual basis and the assumptions required for identification (see especially Todd and Wolpin, 2003, 2007; Rothstein, 2010; Andrabi, Das, Khwaja, and Zajonc, 2011 and Fiorini and Keane, 2014). In this section, I briefly present the empirical framework for the estimation of these models highlighting particularly their implicit assumptions and threats to identification in this particular instance.

4.1 Education Production Functions

The conceptual basis for value-added models is typically derived from a structural cumulative effects model. Following Todd and Wolpin (2003, 2007), it is possible to write the achievement production function in a general form:

$$y_{ist}^* = F[X_i(t), S_i(t), \mu_{is0}, \epsilon_{ist}] \quad (1)$$

where the achievement (y_{ist}^*) of child i in school s at time t is expressed as a function of the whole history of home-based inputs $X_i(t)$, school-based inputs $S_i(t)$, student endowments μ_{is0} (such as ability), and a time-varying error term ϵ_{ist} . While useful for conceptualizing the production technology for achievement, direct estimation of Eq (1) is not typically possible because

the whole history of home and school inputs, as well as individual-specific endowments, are not observed by the researcher.

Equation (1) is typically assumed to be linear, as in the following equation provided by Andrabi, Das, Khwaja and Zajonc (2011):

$$y_{it}^* = \alpha_1' . x_{it} + \alpha_2' . x_{i,t-1} + \dots + \alpha_t' . x_{i1} + \sum_{s=1}^{s=t} \theta_{t+1-s} \mu_{is} \quad (2)$$

where x_{it} is a vector of inputs for child i at time t , y_{it}^* is true achievement at time t measured without error, and summed μ_{is} are cumulative productivity shocks.

Adding and subtracting $\beta y_{i,t-1}^*$ to Equation (1) and assuming that coefficients decline geometrically yields the lagged value-added model:

$$y_{it}^* = \alpha_1' . x_{it} + \beta y_{i,t-1}^* + \mu_{it} \quad (3)$$

Equation (3) is known as the lagged or dynamic OLS (DOLS) VAM. The lagged test score in the above specification is expected to capture the contribution of all previous inputs and any past unobservable endowments and shocks. This is believed to be a significant improvement over a contemporaneous specification, which links current test scores to only current inputs.

Implicitly, however, the specification remains vulnerable to bias from multiple sources, in addition to relying on strong assumptions of geometric decay and linearity as assumed earlier. First, conditioning on the lagged test score introduces the risk of bias due to measurement error which would attenuate the estimate of the persistence parameter (β) but could also bias

the input coefficients (α) in an unknown direction. Secondly, the lagged test score can act as a summary statistic for mental ability only if it enters through a one-time process or endowment. However, if talented children also learn faster, i.e. fixed ability has effects at every age, then β will be upwardly biased (due to positive correlation with ability) and input coefficients may be as well. As Todd and Wolpin (2003, p. F21) discuss, the model then requires two sufficient statistics for identification - the lagged test score and mental ability at baseline: since mental ability may be correlated with the input vector $\mathbf{x}$, and will certainly be correlated with lagged achievement, its omission may substantially bias estimated coefficients.

The interpretation of the coefficients α depends crucially on what the included inputs in $\mathbf{x}$ are. If the input vector includes all contemporaneous inputs then the coefficients in the α vector can be interpreted as technology parameters i.e. the *ceteris paribus* effect of changing one input while keeping all other inputs fixed. This parameter, as Todd and Wolpin (2003) point out, is conceptually different from the parameter identified by randomized field experiments and natural experiments: the latter typically identify a total policy effect which captures the effect of changing the relevant input and allowing all other inputs to respond (i.e. the total derivative with respect to x). In general, the two types of coefficients do not need to agree.

This issue is more generally relevant in the interpretation of results from value-added models. For example, the literature on teacher effectiveness typically estimates teacher effects by regressing current achievement on lagged achievement, a vector of contemporaneous (often fixed) child characteristics and a vector of teacher dummy variables, coefficients on

which are interpreted as teacher value-added. As Todd and Wolpin (2003, p.F27-F28) point out, since these regressions typically do not control for other inputs that could be adjusted in response to teacher quality after the teacher assignment is known (for example, home inputs such as extra tutoring which could compensate for poor teacher quality), they typically identify policy effects and not technological parameters.¹⁴ This distinction will be important in the context of this paper both for interpreting the parameters estimated here and for their comparison to the results from the MS experimental study.

4.2 Are value-added models reliable for the estimation of policy effects?

The primary aim of this paper is to assess if the causal contribution of private schools to learning is significantly greater than the contribution of government schools. This parameter, the differential value-added of private schools, is a policy effect and not the *ceteris paribus* effect of private schooling. In this, my objective is most similar to the literature on identifying teacher and school effectiveness using value-added models.¹⁵

As noted in the previous sub-section, the consistency of estimates from value-added models relies on strong assumptions. It is therefore useful to assess the reliability of value-added models for the identification of treatment effects using results from recent studies. Guarino, Reckase and

¹⁴This difference is independent of any concerns of endogeneity that may arise from some inputs being omitted; for example, these issues of interpretation would apply equally even if the relevant included inputs were randomly assigned.

¹⁵For example, if private schooling results in positive behavioural changes (e.g. an increase in student effort) which result in better test performance, the parameter of interest here should not be estimated conditional on student effort.

Wooldridge (forthcoming) assess the reliability of different value-added models for recovering teacher effects using simulated data with a variety of non-random assignment structures. Specifically, they assess the reliability of six different value-added estimators – a Dynamic OLS specification (as in equation 3), Arellano-Bond panel data estimators, pooled OLS on the gain score, an average residual approach, random effects on the gain score and fixed effects on the gain score – to different combinations of mechanisms to assign students to classrooms and teachers. They report that the DOLS model “was, by far, the most robust estimator across scenarios.”(p.30) and that the “main strength of this estimator lies in the fact that, by including prior achievement on the right-hand side, it controls either directly or indirectly for grouping and assignment mechanisms”.¹⁶ They state that the DOLS estimator, as typically used to estimate teacher effects i.e. not controlling for all home or other inputs, may be thought of as a regression-based version of a dynamic treatment effects estimator (p. 31); it is in this sense that VAMs will be used in this paper.

The Guarino, Reckase and Wooldridge results are based on simulated data. Their relatively positive results mirror patterns in studies that compare effect sizes based on VAMs to quasi-experimental and experimental results in a variety of contexts. Andrabi et. al. (2011) document, while analyzing the effectiveness of private schools in Pakistan (a setting very similar to the one in this paper), that biases from measurement error and unobserved heterogeneity are countervailing, and aggregate bias on the private school coefficient does not seem to be significant.¹⁷

¹⁶References to specific page numbers refer to the Working Paper version of the paper, ?, circulated as an IZA Discussion Paper.

¹⁷Specifically, they report: “Despite ignoring measurement error and unobserved heterogeneity, the lagged value-added model estimated by OLS gives similar results for

Muralidharan and Sundararaman (2013b), in a paper on rural government schools in India, similarly document identical experimental and VA estimates of the effectiveness of contract teachers.¹⁸

Similar results are also emerging from a growing literature in the US: Deming et. al. (forthcoming) compare the effects of a school choice lottery in the US and find no significant differences between experimental estimates of school effects based on the school lottery and estimates from a value-added model that controls for previous test scores; Kane and Staiger (2008) and Kane, McCaffrey, Miller, and Staiger (2013), analyzing results from different experiments that assigned children randomly across classrooms, similarly report that teacher effect estimates that controlled for prior student test scores yielded unbiased predictions of test scores after randomization; Chetty et al. (forthcoming) find no evidence of bias when comparing estimates of teacher effectiveness using a value-added approach to estimates using previously unobserved parent characteristics and a quasi-experimental research design based on changes in teaching staff; and Angrist, Pathak, and Walters (2013) also show how their estimates of Charter school estimates are identical when estimated on the same sample of children using lottery outcomes and separately using observational data (including baseline scores).

the private school effect as our more data intensive dynamic panel methods, although persistence remains overstated. The relative success of the lagged value-added model can be explained by the countervailing heterogeneity and measurement error biases on β (their persistence parameter) and because lagged achievement can also act as a partial proxy for omitted heterogeneity in learning.”

¹⁸Specifically, they show that a reduction in the pupil-teacher ratio has an equivalent impact whether it was caused due to the provision of an additional regular teacher or an experimentally-assigned contract teacher. Combined with additional results presented in their paper, that contract teachers and regular teachers had the same impact on test scores per year, this indicates that VA estimates of the impact of regular teachers were unbiased.

Finally, note that my aim in this paper is different from the teacher effectiveness literature in one important respect – namely, I am not attempting to recover a teacher or school-specific fixed effect. Much of the criticism surrounding VAMs has been regarding how *individual* teacher rankings may be substantially misclassified. For example, Guarino, Reckase and Wooldridge report that even though the DOLS recovers teacher effects reliably on average in most settings, there is considerable scope for the misclassification of individual teachers, thus leading them to caution against their use for high-stakes assessments of individual teachers. This is however, less of a concern for the purposes of this paper since I am not attempting to recover individual school level estimates but only the average difference in value-added between private and government schools.¹⁹

4.3 Estimated specifications

This paper uses a DOLS value-added model as the main specification in the analysis. The estimated specifications are similar across these samples but differ in some details due to different data availability. The core specification used for the estimation in the case of the 8-year old sample is as follows:

$$Y_{it} = \alpha + \beta_1 \cdot Private_{it} + \beta_2 \cdot site_i \quad (4)$$

¹⁹The other key distinction between the literature on teacher effectiveness in the U.S. and this paper is that there are no stakes attached (e.g. performance-based retention or financial rewards related to VA estimates) to the estimates, nor is any such policy move imminent or anticipated. For a discussion of how the addition of financial stakes may compromise the quality of information in VA estimates due to manipulation by teachers or schools, see ?.

$$+\beta_3.X_{i,t} \tag{5}$$

$$+\beta_4.Y_{it-1} \tag{6}$$

$$+\beta_5.timeuse_{it} + \epsilon_{it} \tag{7}$$

where $Private_{it}$ is an indicator variable equaling 1 if the child is enrolled in a private school in 2009/10, with enrolment in a government school as the base category; $site_i$ is a vector of sentinel site (mandal) dummy variables; X is a vector of background characteristics that includes caste dummies and wealth index, maternal and paternal years of schooling, household size, the sex of the child and whether he/she is the first-born child; $Y_{i,t-1}$ is the lagged test score; $timeuse_{it}$ is the number of hours spent on a typical day in various activities - specifically, I control for the time use on caring for others, domestic tasks, studying outside of school time (including extra tuition), tasks on the family farm or other family business, time spent in school, time spent in play or general leisure, and paid work outside of the household (time sleeping is the omitted category).

The estimation includes controls sequentially as in Equations 4-7 and therefore the interpretation of β_1 , the parameter of interest, changes with the inclusion of further controls. Equation (4) measures the unconditional cross-sectional difference in test scores within-community between students from private schools and government schools; Equation (5) presents a contemporaneous specification of the value-added model which controls additionally for a vector of background characteristics; Equation (6) is the Dynamic OLS specification of the lagged VAM; Equation (7) additionally includes time use.

Following the interpretation of Guarino, Reckase and Wooldridge (2012), β_1 in equations 4-6 can be thought of as a treatment effect, with the sequential addition of controls bolstering the reliability of estimates. Interpretation of equation (7) is not as clear immediately. In adding a full vector of time use categories, I follow the lead of Fiorini and Keane (2014) who similarly include a vector of time use categories in studying skill formation. However, while the inclusion of a vector of time use inputs carries potential advantages, it also presents challenges for interpretation and identification. Time spent in different activities, such as studying, is clearly an input into the production of academic achievement, the omission of which could lead to bias in the estimation of the differential value-added by private schools. For example if, as is consistent with our data, students who go to private school studied longer after school and worked less on domestic tasks, then any differences in academic achievement caused by this differential allocation would be captured in coefficient β_1 in equation (6). This is not a problem for the identification of a treatment effect if the differences in time allocation were a result of private school attendance caused, for example, by private schools assigning more homework; in that case, differential time use is merely the channel through which private school effects are realized and these should not be controlled for. If, however, differential time use is a result of unobserved student background characteristics (e.g. parental motivation) then its omission will bias the resulting estimate of β_1 in Equation (6); that this may be a plausible scenario is indicated by the results of the MS study who document that time studying at home did not adjust upwards for lottery-winners who went to private schools. Given these issues surrounding the inclusion of time use inputs, I present estimates

of β_1 both from equation (6) and equation (7).²⁰ The inclusion of time use inputs is expected to provide lower-bound estimates of the treatment effect of private schools since it effectively assumes that the type of school does not affect time allocation patterns; conversely, omission of the time use vector provides an upper bound impact.²¹

In the 15-year sample, specifications are identical to the 8-year olds with one exception. Since about a fifth of the sample is no longer enrolled in school, I also include a dummy variable for not being enrolled in school in 2009/10.²² In the 9-year old sample (2011), time use data was not collected and I use the time use as reported at 8 years (2010) as proxies for time allocation in 2011.²³

²⁰The inclusion of time allocation patterns in value-added models also implicitly imposes a further restriction, namely that current productivity shocks do not result in (or, more generally, are correlated with) a change in time allocation. This is a strong assumption but one that is commonly made in the literature for example in Fiorini and Keane (2014) but also implicitly in Todd and Wolpin (2007) where they include the HOME score which includes various home investments such as the frequency of reading to a child and conducting various activities with the child; in both cases, the restriction that current productivity shocks do not affect child time use or parental time investments is assumed.

²¹This is not an assertion that can be made in general; for example, a priori it is entirely possible that parents compensate for lower quality government schools by making children study more at home, in which case the inclusion of time use inputs would bias upwards the total policy effect of private schools. It is made in this specific instance because of a noticeable pattern in the data where children in private schools spend significantly more time studying and because the MS study does not find differences in time spent studying at home for lottery-winners: if parents were indeed compensating for private school attendance by making children study more after school, we should have seen time studying at home to reduce significantly for the lottery-winners who attended private schools.

²²The identifying assumption for the differential value-added is still one of conditional exogeneity, namely that conditional on covariates and lagged achievement assignment across type of school (including being not enrolled) is no longer correlated with the error term.

²³Note that if both TU_t and TU_{t-1} were observed and included in the estimation, any significant effects of TU_{t-1} would indicate that the structural assumptions underlying the lagged value-added model are violated. If only TU_{t-1} is included, given that it is expected to proxy for current time use, the significance of lagged time use does not provide any clear evidence of such violation.

Using lagged time use to proxy for current time use implicitly imposes a strong assumption that agents do not re-optimize time use patterns upon the realization of

In the 8-year sample, I use scores from the maths test and the PPVT as outcome variables. I use the 2007 score on the quantitative section of the Cognitive Development Assessment as the lagged score for math and the 2007 score on the PPVT as the lagged measure for the PPVT in 2009/10. In the 9-year old sample, I have three test measures: a maths test (which had common items with the test administered in 2010), a test of Telugu competence and a test on English language competence.²⁴ I use the maths test in 2010 as the lagged measure for 2011 math test and the PPVT (administered in Telugu) as the lag for the Telugu test. Since a test in English was administered for the first time in 2011, I use the PPVT receptive vocabulary test score as the lagged achievement measure for English. In the 15-year old sample, I use PPVT, maths and Telugu as the outcome variables.²⁵

In the main body of the paper, I only report results from Equations (6) and (7) but results in detail at each step are presented in Appendix tables A1 to A3 for rural areas. In all regressions in this paper, I cluster standard errors at the sub-district (mandal) level.²⁶

current period shocks; it is, however, implied in the restriction of current time allocation being independent of the error term which is made in going from Equation (6) to Equation (7).

²⁴At this point, children in the younger cohort were aged about 9-10 years which is exactly analogous to the age of the children in the MS study four years after their experimental intervention offering scholarships. Furthermore, they test the children on all of the three dimensions (Math, Telugu and English) in which test scores are available to us. Therefore results on this sub-sample are the most comparable to their experimental estimates.

²⁵There is a difference in the period between the lagged assessment and the outcome assessment across the three age samples which makes the estimates across the samples incomparable as the dosage of treatment is different. This is addressed more fully in Section 5.

²⁶Sub-districts, used for generating site fixed effects and for clustering standard errors, are taken as the first sub-district that the child was observed in (i.e. in 2002). For a subset of children whose households had moved out of the initial communities, but were tracked to their new location for the data collection, this is not the same as the sub-district they were observed in 2010. The reason for this choice is that the locations

Table 1: Descriptive statistics - Younger Cohort, 8 years (2010)

	Rural Areas			Urban Areas		
	Government	Private	Difference	Government	Private	Difference
Mother's education (years)	3.06	4.48	-1.42***	3.72	7.86	-4.13***
Father's Education (years)	4.72	7.21	-2.49***	4.97	9.10	-4.13***
Male	0.47	0.65	-0.18***	0.47	0.57	-0.10
First-born child	0.32	0.49	-0.16***	0.25	0.46	-0.21***
Scheduled Caste	0.24	0.15	0.09***	0.21	0.08	0.13**
Scheduled Tribe	0.19	0.11	0.08***	0.07	0.01	0.06
Other Backward Classes	0.49	0.49	-0.00	0.39	0.47	-0.07
Other castes	0.09	0.25	-0.17***	0.33	0.44	-0.11
Household size (2010)	5.53	5.52	0.00	5.43	5.16	0.27
Monthly per capita expenditure (2006)	622.80	931.58	-308.79***	699.15	1102.24	-403.09***
Monthly per capita expenditure (2010)	900.77	1452.61	-551.83***	1102.00	1740.78	-638.78***
<i>Time use (hours spent on 'typical' day)</i>						
caring for others	0.22	0.20	0.03	0.11	0.17	-0.07
household chores	0.40	0.27	0.13***	0.26	0.25	0.01
at school	7.56	8.03	-0.47***	7.59	7.94	-0.35**
studying after school	1.76	2.01	-0.26***	1.63	1.96	-0.32*
unpaid work outside household	0.02	0.00	0.01	0.00	0.00	0.00
paid work outside household	0.01	0.00	0.00	0.01	0.00	0.01
play/general leisure	4.85	4.32	0.53***	5.21	4.62	0.59**
sleeping	9.17	9.16	0.02	9.17	9.02	0.15
N	956	476		76	382	

Cells present mean of variable within school type. Asterisks indicate significant difference between mean in private and government schools within urban and rural areas. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.4 Results

Descriptive statistics about the background characteristics of children in the younger and older cohort, disaggregated by the type of school in which they are enrolled, are provided in Tables 1 and 2.

In both rural and urban areas, there are significant differences in the observable characteristics of children in government and private schools: children in private schools are likely to be from richer households with more educated parents and are much more likely to be male and first-born children. The share of private school enrolment varies across cohorts and of the migrated children is not available in the public-use Young Lives dataset other than a new category to indicate that children had migrated. Results are unchanged when restricted to the children whose households are still in the same communities as originally surveyed in 2002.

Table 2: Descriptive statistics - Older Cohort, 15 years (2010)

	Rural Areas			Urban Areas		
	Not enrolled	Public School	Private School	Not enrolled	Public School	Private School
Mother's education (years)	1.27***	4.13	4.44	3.42	4.06	7.09***
Father's Education (years)	2.49***	4.67	6.72***	4.77	5.91	9.54***
Male	0.41	0.49	0.58	0.48	0.39	0.58**
First-born child	0.23	0.26	0.43***	0.23	0.33	0.48*
Raven's test score (2002)	-0.10	-0.09	0.10	0.07	0.01	0.31
Scheduled Caste	0.22	0.27	0.11***	0.23	0.22	0.05**
Scheduled Tribe	0.12	0.13	0.10	0.00	0.03	0.01
Other Backward Classes	0.55	0.48	0.47	0.32	0.52	0.49
Other castes	0.10	0.12	0.32***	0.45*	0.21	0.44***
Household size (2010)	5.30	5.20	5.12	4.94	5.13	5.15
Monthly per capita expenditure (2006)	721.94	795.61	1095.98***	936.70	869.81	1308.50***
Monthly per capita expenditure (2010)	1308.15	1179.14	1504.46**	1276.80	1471.84	2111.73**
Time Use (hours on typical day)						
caring for others	0.58***	0.21	0.18	0.42	0.19	0.14
household chores	2.47***	1.44	0.86***	2.10**	0.96	0.76
at school	0.39***	7.87	8.58***	0.00***	7.70	8.81***
studying after school	0.16***	2.40	2.83**	0.42***	2.66	2.60
unpaid work outside household	1.95***	0.18	0.13	0.58	0.00	0.04*
paid work outside household	4.53***	0.06	0.00**	4.71***	0.00	0.00
play/general leisure	5.25***	3.66	3.47	7.29***	4.54	3.49**
sleeping	8.70***	8.21	7.96**	8.61*	8.22	8.19
N	189	408	137	31	67	136

Cells present mean of variable within school type. Asterisks indicate significant difference between mean in private and government schools within urban and rural areas. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Cross-sectional differences in test scores (Private - Government)

	Math	PPVT	Telugu	English
Rural				
8 years (Younger cohort, 2010)	0.32***	0.36***		
9 years (Younger cohort, 2011)	0.49***		0.25***	1.05***
15 years (Older cohort, 2010)	0.49***	0.67***	0.45***	
Urban				
8 years (Younger cohort, 2010)	0.45***	0.25*		
15 years (Older cohort, 2011)	0.34***	0.27**	0.18	

Cells show cross-sectional mean difference in test scores. Asterisks indicate if difference is significantly different from zero (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). Test scores are generated using Item Response Theory models and internally normalized (see Sec. 2.2). Only 23 children are in urban government schools in 9-year old sample. Results for this group are not reported.

across urban and rural areas. Furthermore, whereas nearly all children in the younger cohort are enrolled in school, by the age of 15 about 22 per cent of children in this sample have dropped out of schooling.

Table 3 presents the raw magnitudes of the test score gaps between children in private and government schools across the three samples and across urban and rural areas.²⁷

In Table 4, I present the coefficient on the private school dummy from Equation (7), which includes all available controls, separately for rural areas, in all three samples.

In rural areas (Panel A), at the age of 8, there is a significant private school premium of about 0.13 SD in PPVT which is about 35 per cent of the size of the within-community raw premium in test scores; there is no significant

²⁷Since the 9-year old sample has only 23 children in urban government schools, I do not report any results for children in urban areas for this sample. The very small sample size restricts me from making any reliable conclusions about results in this sample for urban areas.

premium in maths test scores. At the age of 9, there is no significant private school premium in maths but there is a very large private school premium in English of about 0.65 SD; in Telugu, there appears to be a negative effect of attending private schools; this cohort is similar to the sample in MS and the pattern of incidence of private school effects is also similar.

Table 4: Coefficient on Private School dummy

Rural Areas									
	Math	Without time use			Math	With time use			N
		PPVT	Telugu	English		PPVT	Telugu	English	
8 years (YC, 2010)	0.018	0.17***			-	0.13**			1,435
		(0.064)			0.063	(0.059)			
9 years (YC, 2011)	0.091		-0.18	0.73***	0.058		-	0.66***	767
		(0.056)	(0.10)	(0.15)	(0.060)		0.23**	(0.14)	
15 years (OC, 2010)	0.20***	0.18*	0.10		0.20**	0.16*	0.14		729
		(0.065)	(0.085)	(0.074)	(0.075)	(0.088)	(0.080)		
Urban Areas									
	Math	Without time use			Math	With time use			N
		PPVT	Telugu	English		PPVT	Telugu	English	
8 years (YC, 2010)	0.030	0.058			-	0.084			458
		(0.15)	(0.24)		0.028	(0.16)	(0.21)		
15 years (OC, 2010)	0.10	-0.019	-0.023		0.092	-0.062	-0.11		233
		(0.14)	(0.050)	(0.21)	(0.11)	(0.065)	(0.19)		

YC = Younger cohort, OC= Older cohort; *** p<0.01, ** p<0.05, * p<0.1
Cells correspond to the coefficient on the Private school dummy with government schools as the base category. Regressions control for child and household characteristics, sub-district fixed effects, lagged achievement score (Eqs. 6 and 7). See Sec 3.2 for details. Coefficients of other variables, and more parsimonious specifications (Eq:4-6) are presented on Appendixes A.1 to A.3 for rural areas. Standard errors, in parentheses, are clustered at sub-district (mandal level). N refers to number of observations in the regression on math test scores.

At the age of 15 years, there is a significant private school premium in all test scores (although not statistically significant for Telugu). The size of

the premium in this cohort is relatively modest: about 0.15 SD for PPVT and the Telugu cloze test and about 0.2 SD for the math test; in each case, this is between 20–40 per cent of the raw baseline difference.

In urban areas, as is evident from Panel B, I do not find any statistically significant evidence of a private school premium in either cohort. While this could potentially be due to relatively small sample sizes in urban areas, this does not seem to be the case in practice: in the 15-year old sample, the coefficient itself is very close to zero; and the problem of low sample size is not as severe a problem for the 8-year-old sample as it has twice the number of observations. Certainly these results suggest that even if there is a private school premium in urban areas, which is not detected due to relatively small sample sizes, it is very unlikely to be large in magnitude.

As pointed out by previous research, as well as the MS study, the private school sector is very heterogeneous. A key aspect of heterogeneity is the medium of instruction; while government schools teach all subjects in the local language, private schools may either use Telugu or English as the medium of instruction.

In Table 5, I present results which re-estimate regressions for the 9-year old sample, distinguishing between English-medium and Telugu-medium private schools (with government schools as the base category).

There is clear evidence of a private school premium in English across both Telugu and English medium private schools with the magnitude (about 0.8 SD) being expectedly greater in the latter but still substantial at about 0.5 SD even in Telugu-medium private schools. Importantly, the negative effect of private schools on Telugu seems to be concentrated entirely in the English-medium private schools and although substantial at 0.35 SD,

it is still much smaller than the positive premium of 0.8 SD in English.²⁸ A further interesting pattern to note is that the coefficient on going to an English-medium private school, while not statistically significantly different from zero, is negative for mathematics while that for a Telugu-medium private school is positive; this also corresponds closely with patterns documented by MS.²⁹

Table 5: Heterogeneity by medium of instruction - Rural Areas, 9 years

	Without time use			With time use		
	Math	Telugu	English	Math	Telugu	English
Private School - English medium	-0.053 (0.085)	-0.35** (0.15)	0.83*** (0.14)	-0.095 (0.079)	-0.39** (0.15)	0.77*** (0.13)
Private school - Telugu medium	0.24* (0.13)	-0.032 (0.12)	0.64*** (0.15)	0.22 (0.13)	-0.090 (0.11)	0.55*** (0.15)
F-statistic (equality of coefficients)	3.26	5.60	3.65	4.03	5.46	4.43
Prob >F	0.092	0.033	0.077	0.064	0.035	0.054
Observations	766	708	670	766	708	670

Cells correspond to the coefficient on the dummy variables for type of private school with government schools as the base category. Regressions control for child and household characteristics, sub-district fixed effects, lagged achievement score (Eq:7). See Sec 3.2 for details. Standard errors, in parentheses, are clustered at sub-district (mandal level).*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

While the pattern of incidence is identical, *prima facie*, there appears to be a major difference in the size of the private school effect in English; in their main results, relying on the random allocation of vouchers for identification,

²⁸In the absence of estimates of labour market returns to English and Telugu, it is not clear how these trade-offs should be weighted (a point also made by MS). However, it is reasonable to assume that returns to English are significantly greater than local languages including Telugu - this is, for, example, the pattern documented by Munshi and Rosenzweig (2006) in Mumbai - and certainly seems to be the impression among parents, who view the additional English language proficiency provided by private schools as one of their biggest draws.

²⁹The coefficients on English and Telugu-medium private schools are however always significantly different from each other, in Telugu at 5% level of significance and in English and Mathematics at 10% level of significance.

MS report a Local Average Treatment Effect (LATE) effect size of about 0.32 SD at 2.5 years (and even smaller at just above 0.2 SD after four years) which is half the size of the effect I find.

This discrepancy in results could be caused due to various reasons. The first possible reason relates to the interpretation of the LATE parameter as estimated by the MS study. If the effect of private schooling is heterogeneous then, even in the absence of bias, treatment effects estimated will differ depending on the choice of IVs due to the change in the population of ‘compliers’ over whom the treatment effect is estimated. That this may partly be the case is highlighted by the results on heterogeneity by medium of instruction that MS present. In those regressions, where medium of school is instrumented using an interaction of receiving a voucher and the medium of instruction of the nearest private school, the treatment effect at 2.5 years is about 0.53 SD for English medium schools and 0.4 SD in Telugu-medium schools.

A further possibility could arise from the choice of test scores used in this study and in the MS study: whereas I use IRT scores in this paper, MS use a standardized raw score as their dependent variable. Since Item Response models allow items to differ in difficulty and other characteristics (e.g. guessing), they may change the contribution of each test item towards a composite test score and thereby the spread of the distribution. I reran the estimation using a (normalized) raw score as the dependent variable, following the procedure in MS. The main effect size I get is 0.32 SD which is nearly identical to the estimate reported by MS.

That the choice of aggregation method should create such large changes in effect sizes raises questions about the reliability of the results. As a further

Table 6: Performance on specific English tasks

VARIABLES	(1) Match number to word	(2) Match picture to word	(3) Match sentence to word	(4) Complete sentence (MCQ)	(5) Paragraph comprehension questions	(6) Complete sentence (open-ended)	(7) Write sentence description
Private School	0.25*** (0.069)	0.26*** (0.064)	0.18*** (0.045)	0.082** (0.032)	0.067** (0.026)	0.038** (0.016)	0.15*** (0.029)
Govt. school mean	0.63	0.68	0.56	0.39	0.23	0.02	0.03
Observations	670	670	670	670	670	670	670
R-squared	0.316	0.313	0.210	0.238	0.186	0.571	0.330

Cells correspond to the coefficient on the Private school dummy with government schools as the base category. Regressions control for child and household characteristics, sub-district fixed effects, lagged achievement score (Eq:7). See Sec 3.2 for details. Standard errors, in parentheses, are clustered at sub-district (mandal level).*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

check, to ensure that results reported here are meaningful, I calculated the proportion correctly answered on the English test for each different task administered and used them as separate dependent variables. The test administered seven distinct sets of tasks: linking numbers to words, linking pictures to words, linking pictures to sentences, completing a sentence by choosing a word from one of three options, answering questions based on reading a passage, open-ended cloze ('fill-in-the-blanks') items for sentence completion, and writing a sentence to describe a provided picture. Results are reported in Table 6 along with the mean proportion correctly answered by the government school students. As can be seen, the private school effect is significantly positive across all specifications indicating that the existence of a private school effect on English is not an artefact of the procedure undertaken to create a test score.

Moreover, expressing the private school effect as a percentage of the mean proportion correct in the government sector, the effects are larger on harder

questions such as writing a complete sentence to describe a picture than on easier items such as linking numbers or pictures to words. The discrepancy between effect sizes could be a product of the implicit reweighting of items in the IRT scores and suggests that weighting all test questions equally (as the creation of raw scores does) understates the effect of private schools.

The close correspondence between patterns in this paper and in the MS study is particularly striking. The nearly identical estimates indicate, in addition to a lack of bias in the VA estimates of the private school effect, that schools have not made material adjustments to their inputs as a result of the MS voucher and that there has also not been a large adaptation on the part of parents/households. MS document that lottery winners did not, in fact, change their time use patterns; further it seems unlikely that a one-off intervention applying only to one cohort of children (in kindergarten and Grade 1 at the start of the experiment) would lead schools to adapt their long-term production strategies.³⁰

The results still have an unresolved interpretational issue. In the 8 and 15 year samples, the treatment effects estimated at best identify three-year effects of private school enrolment which, while still adequate to compare the government and private sectors within the samples, are not comparable to results for the 9-year old cohort (since the dosage there is one year).

³⁰Put differently, the MS study differs from other voucher experiments in developing countries (see e.g. Hsieh and Urquiola, 2006 with data from Chile) since schools have not exhibited the full set of general equilibrium effects that might be expected if the voucher system were general, long-term and large in scope (as may indeed happen with the Right to Education Act in India). This, combined with the low evidence of household adaptation of inputs in response to the randomized receipt of vouchers, renders their results comparable with the VA estimates here.

For an example, using both experimental and panel-based variation from different settings, of how household adaptation to interventions can change policy effects see Das, Dercon, Habyarimana, Krishnan, Muralidharan, and Sundararaman (2013).

Under the assumptions of constant treatment effect for each year of schooling in each cohort of children, and geometric decay, this effect may be calculated if the true persistence parameter is known. Recovering the ‘true’ persistence parameter however, as Andrabi et. al. (2011) highlight in a similar application, requires substantially more stringent assumptions and richer data; in particular, the persistence parameter estimated in the dynamic OLS VAM is likely to be biased. Andrabi et. al. (2011) document persistence parameters of between 0.2 and 0.5 i.e between one-fifth and one-half of learning persisting from year to year.³¹ Using these persistence parameter estimates, I can calculate the implied per-year effect from the three-year effects for the 8 and 15 year samples in Table 4. Assuming persistence of 0.2 per year implies, in the 8-year old sample, a per-year effect of about 0.1 SD (~ 0.075 SD if persistence is 0.5) and at 15 years, a per-year effect size of 0.15 SD in math and 0.12 SD in (~ 0.1 SD in math and 0.08 SD in PPVT if persistence is 0.5). These effect sizes appear modest at best.³²

³¹The implied persistence coefficient in other South Asian studies is sometimes even lower. For example, the two interventions in Banerjee, Cole, Duflo, and Linden (2007) show rapid fade-out and lead to implied persistence between .086 and 0.265 in math and below 0.06 for verbal abilities. Imperfect persistence has also been documented in various US datasets (e.g. Jacob, Lefgren, and Sims, 2010, Rothstein, 2010 and Kane and Staiger, 2008) and seems to be a feature of education data more generally.

The MS study documents smaller effect sizes of private schooling on English after 4 years of private school attendance compared to 2.5 years, although not significantly so. Note also that our estimates at 9-years are actually one-year estimates while MS results are two-year effects; that the two sets of results happen to overlap almost exactly, is consistent with low persistence (but also with other possible explanations such as private school advantage in test scores being built at school entry ages and then persisting). While this does not necessarily indicate zero persistence, it does highlight the pitfalls of using short-term teacher or school evaluations, whether randomized or non-experimental, to make assessments of long-term effects.

³²This is not surprising given that even the three-year impacts appear quite modest: any persistence will cause one-year impacts to be less than the three-year impacts with the latter being upper-bound estimates of what one-year impacts may be.

5 Additional concerns and robustness

While previous studies on the robustness of VAMs have been encouraging, and indeed results in this paper on a comparable cohort and indicators conform closely with comparable experimental evidence, the possibility of bias in the estimates cannot be definitively ruled out; this may especially be a concern for indicators/cohorts for which external validation through the MS study is not available. From the perspective of estimating a treatment effect, a particularly important concern relates to potentially unobserved heterogeneity that may still bias the private school coefficient. Bias could also be caused due to dynamic sorting across school types. In this subsection, I evaluate the empirical relevance of these concerns.

5.1 Unobserved differences in ability

Lagged achievement suffices to account for ability differences between children only if these differences affect initial conditions and not learning at each point in time; as previously noted, in the latter case two sufficient statistics are required namely past achievement and the fixed ability measure. The richness of the Young Lives data allow me to investigate the severity of this bias directly for the older cohort of children, aged 15 in 2009.

In the first round of data collection in 2002, when the older cohort of children were 8 years old, the survey had administered the Raven's colored progressive matrices test which is a test of non-verbal reasoning which has

Table 7: Assessing bias due to omitted ability

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Test scores (2009/10) Math	PPVT	Telugu		Private school (2009/10)		Test scores (2006/7) Math	PPVT
Private school	0.19** (0.075)	0.15 (0.086)	0.13 (0.076)					
Raven's score (2002)	0.084*** (0.025)	0.11** (0.048)	0.082*** (0.022)	0.060** (0.022)	0.016 (0.018)	0.010 (0.020)	0.12*** (0.035)	0.15*** (0.035)
Math score (2007)	0.42*** (0.048)				0.037* (0.020)			
PPVT score (2007)		0.64*** (0.067)	0.48*** (0.062)			0.077** (0.027)		
Controls	Y	Y	Y		Y	Y	Y	Y
Time use	Y	Y	Y		Y	Y	Y	Y
Site fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	729	715	706	545	544	544	729	728
R-squared	0.549	0.639	0.519	0.122	0.293	0.303	0.338	0.468

Columns 1-3 regress test scores on full set of covariates and the Raven's test score; coefficients on private school should be compared to Table 4. Columns 4-6 investigate partial correlations between the Raven's test and private schooling, for children in school in 2009/10, with increasing sets of controls. Columns 7-8 investigate partial correlations between lagged test scores (from 2006/7) and the Raven's test. Controls are defined as in main regression specification (Sec. 3.3). Standard errors are clustered at the site level.

widely been used as a proxy for intelligence or innate ability.³³ In the Indian case, it has been used as such by Kingdon (1996) and Kingdon and Teal (2007, 2010). In the context of our application, it provides a possible way to test the structural assumptions of the dynamic VAM.

Given the assumptions of the VAM, if lagged achievement is a summary statistic for fixed ability and for all previous investments into cognitive ability, then the addition of the Raven's score from 2002 should not have any significant effect on achievement at 15 (2009/10) when conditioning

³³For instance, the Raven's test has been used as such by Boissiere, Knight, and Sabot (1985), Alderman, Behrman, Ross, and Sabot (1996a,b), Glewwe (1996), Heady (2003) and Akresh, Bagby, De Walque, and Kazianga (2012). The Raven's test score is thought to be less affected by schooling than most other measures of cognitive ability although it has been shown to be affected by early childhood investments such as in nutrition.

for achievement at age 12 (2007). Columns 1-3 in Table 7 show the effect of adding the Raven's score into the VAM used previously for the 15 year old sample. As can be seen, the Raven's score is significantly associated with learning at 15, even when conditioning on achievement at 12. To the extent that the Raven's test is proxying for innate ability, or at least ability pre-determined at age 8, this pattern suggests that ability does not merely enter as a one-off endowment but has effects in every period. The second pattern to note, however, is that the private school coefficient has barely changed and is statistically indistinguishable from the result in Table 4. Given that the significant coefficient on the Raven's test score indicates a violation of the VAM assumption, this may appear surprising.

An explanation of this pattern can be discerned by looking at the partial correlations between private schooling, lagged achievement and the Raven's score. Specifically, I first regress private school attendance on the Raven's test and then add the background set of controls and lagged test scores from 2007 into the equation.³⁴ As may be noted, while not controlling for background characteristics and lagged achievement, there is a weak positive association between the Raven's score and private school enrolment; upon inclusion of background characteristics and the lagged achievement measure(s), this association is no longer significant and with a coefficient very close to zero indicating that the lagged achievement and background controls do in fact proxy for any selection by ability.

Given that the Raven's score at 8 is correlated with test scores at 12, the possibility of a biased treatment effect remains since any bias in the coefficient on lagged achievement caused due to the omission of the Raven's

³⁴This regression is restricted to children who were observed in school in 2009/10.

score could transmit to the private school coefficient. The magnitude of any such bias on the private school coefficient depends multiplicatively on the partial correlation between private schooling and the lagged test score, and the partial correlation between the lagged test score and the ability measure (Raven's score). In this instance, both objects are quantitatively small – a 1 SD change in the Raven's test at 8 (which would be a very large change) has an effect of about .15 SD on the test score at age 12; in turn, a 1 SD change in the lagged test score has an effect of about 0.04 - 0.07 on the probability of private school enrolment. Taken together, the modest partial correlations account for why the omission of the Raven's test does not significantly change the estimates of the private school effect.

Finally note that, given the positive association of the Raven's test with private school enrolment, these results imply that any bias from the omission of ability measures overstates the private school effect. It does not therefore invalidate the major conclusion of this paper that private school effects are at best modest in most dimensions of learning with the exception of English.

5.2 Parental information, aspirations and dynamic sorting

The robustness check above suggests that the lagged test score suffices to control for sorting by ability (at least as measured by the Raven's test) into private schools. However, a further related concern is that of dynamic tracking. Analogous to Rothstein's (2010) criticism (delivered in the context of tracking of students into different classrooms by headteachers),

while VAMs may deliver unbiased estimates of the private school effect if selection was only on the variables controlled for and past achievement, it is plausible that parents observe more or different information on child achievement which is used as basis for selecting whether the child is enrolled into private or government schools.³⁵

Parent’s assessments of the child’s academic performance may contain information other than that contained in test scores for at least two reasons. Parents may observe much more about their children than our survey measures can capture; and parental assessments may have significant measurement error of their own (if, for example, parents cannot reliably assess a child’s actual progress i.e. how well the child *should* have done as opposed to actual achievement). Furthermore, it is always possible that parents differ in their degrees of aspirations for children and the preferences they have towards their education; if these preferences lead to a greater propensity to select into private schools (as they are perceived to be of higher quality) and also lead to higher home-based investment which is not captured in our range of controls or proxied by past achievement, then our estimate of the private school effect might be biased. The precise reason for (possible) divergence of parental assessments from achievement data

³⁵Rothstein’s (2010) key contribution was to develop falsification tests assessing whether future teacher assignments had an ‘effect’ on current test scores in a value-added framework; he finds, using data from North Carolina that they do, which indicates dynamic tracking not adequately controlled for in the value-added specification. Although later research has sometimes contested the implications of these findings for the reliability of VAMs, e.g. by showing that the falsification test is not violated in Rothstein’s data when correcting for small sample estimates (Kinsler, 2012) or by demonstrating that even otherwise, dynamic sorting and a rejection of the Rothstein test need not necessarily imply biased teacher estimates (Goldhaber and Chaplin, 2012), that is not my major concern here. Sorting which is not controlled for by the lagged achievement measures and controls *could* be a potential source of bias in my results; my focus is to assess, to the degree possible, the magnitude of such bias for my substantive findings about the relative effectiveness of private schools.

on our test measures is not central to the issue; what is important is that selection on ability, if any, depends on the former (parental) measure and not the latter (test scores). If there is systematic divergence between the two, it is plausible that bias may still exist.³⁶

I attempt to quantify the magnitude of bias that may exist due to these sources of bias by using unique proxies available in the Young Lives data for these sources of bias. In 2007 and 2009/10, in both cohorts, the household survey collected parents' assessments of how they thought the child (if enrolled at the time) was performing in school; the measure was collected on a five-point scale with 1 being "Excellent" and 5 being "Very bad". Furthermore, in 2007 the survey asked parents what they would desire as the highest level of education for their child, in the absence of any constraints. These measures appear meaningful: average test scores in mathematics seem to increase incrementally for each point of the parental assessment scale; similarly, parental aspirations about a child's education (reduced to a dummy variable for whether the parent would like the child to go to university) are positively associated in bivariate correlation with private school attendance. As a robustness check on this possible source of bias, I estimate the lagged VAMs on the all three age samples supplementing the specification with a vector of dummy variables for each

³⁶I attempted to test directly the hypothesis that selection into private schools based on academic performance is driven exclusively by parental assessments of academic performance and not the objective test measures of performance. Specifically, I first regressed private schooling on the household controls and lagged test scores for each sample and note that lagged test measures are correlated with later private school attendance. I then add parental assessment of academic performance and parental aspirations for child education. I note that whereas these parental measures show correlations with private schooling that are statistically significant and quantitatively large in the expected direction, the coefficient on lagged test score is no longer significantly predictive of later private school enrolment. This provides suggestive evidence that the posited intuition about determinants of tracking is not misguided. Results are available on request.

Table 8: Robustness to selection on parental assessments and aspirations - Rural areas

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	8-year old (2010)		9-year old (2011)			15-year old (2010)		
	Math	PPVT	Math	Telugu	English	Math	PPVT	Telugu
Private	-0.11*	0.12*	0.039	-0.33**	0.64***	0.21***	0.21***	0.14**
	(0.058)	(0.062)	(0.080)	(0.12)	(0.13)	(0.061)	(0.068)	(0.060)
Performance: Good	-0.12	0.093	-0.17	-0.40*	-0.33***	-0.15	-0.11	-0.069
	(0.10)	(0.11)	(0.20)	(0.20)	(0.077)	(0.093)	(0.085)	(0.077)
Performance: Reasonably Well	-0.23*	-0.027	-0.21	-0.54***	-0.41***	-0.40***	-0.085	-0.19
	(0.11)	(0.12)	(0.21)	(0.18)	(0.079)	(0.12)	(0.093)	(0.12)
Performance: Poorly	-0.35**	-0.25	-0.28	-0.39*	-0.40*	-0.66***	-0.17	-0.71***
	(0.13)	(0.15)	(0.24)	(0.21)	(0.22)	(0.19)	(0.12)	(0.19)
Performance: Very Bad	-0.89***	-0.69**	-0.17	-0.82	-0.29	-1.22	-0.37**	-1.92***
	(0.17)	(0.30)	(0.41)	(0.65)	(0.54)	(0.89)	(0.17)	(0.19)
Parent Aspiration: Child will go to university	0.11*	0.033	-0.017	0.025	0.023	0.015	0.071	0.065*
	(0.053)	(0.041)	(0.091)	(0.076)	(0.064)	(0.054)	(0.061)	(0.034)
Observations	1,204	1,068	648	597	570	620	607	603
R-squared	0.364	0.327	0.515	0.367	0.419	0.410	0.593	0.433

Standard errors clustered at sub-district (mandal) level. *** p<0.01, ** p<0.05, * p<0.1 Regressions control for child and household characteristics, sub-district fixed effects and lagged achievement score. Performance: 'Excellent' is the omitted category.

point of the parental assessment scale (with “Excellent” as the omitted category) and a dummy variable for whether the parent desires the child to go to university.³⁷

Results from this analysis are given for rural areas in Table 8; as can be seen, even though there is information in the parental assessments and their educational aspirations which is related to test scores, the coefficients on the private school dummy variable seem to be unchanged from the main estimates.³⁸ I find no evidence of additional bias in the VAM specification estimated in the previous subsection.

5.3 Dosage of treatment across samples and school migration

The main results documented in Section 4 for children in the 8 and 15 year old samples rely on lagged achievement from three years ago (2006) and the type of school attended in 2009. This approach ignores any switching of schools between rounds which can contribute to bias.

The 2009 round of the survey collected retrospective school histories of children, eliciting their enrolment in various schools starting from the 2009/10 school year and going backwards. As James and Woodhead (2014) document, there is a considerable degree of mobility across schools in these

³⁷In 2007, children in the 8-year cohort were aged between 4.5–6 years and only about 44 per cent had joined formal schooling; for enrolled children parental assessment of performance at school was collected. Most other children were in preschools (including *anganwadis*) and the survey asked for the parent’s assessment of child performance there. Together these two variables allow me to construct lagged measures of parental assesment for the 8-year old cohort.

³⁸This is true for most coefficients in the regressions apart from the lagged achievement measures which decline in magnitude. This indicates that parents’ assessments of child performance, although informative, do not seem to bias the estimates and probably reflect information similar to the lagged achievement measures.

Table 9: Years spent in private and government schools between 2007/8 and 2009/10 academic years (inclusive)

8 years old													
Rural areas							Urban areas						
Years in private schools							Years in private schools						
	0	1	2	3	Total		0	1	2	3	Total		
Years	0	0	11	38	332	381	Years	0	0	0	4	369	373
in	1	49	9	60	0	118	in	1	3	2	4	0	9
govt.	2	180	51	0	0	231	govt.	2	15	11	0	0	26
schools	3	702	0	0	0	702	schools	3	50	0	0	0	50
Total	933	71	98	332	1,434		Total	68	13	8	369	458	

15 years old													
Rural areas							Urban areas						
Years in private schools							Years in private schools						
	0	1	2	3	Total		0	1	2	3	Total		
Years	0	114	4	7	97	222	Years	0	16	3	4	131	154
in	1	17	2	22	0	41	in	1	3	1	3	0	7
govt.	2	48	29	0	0	77	govt.	2	5	7	0	0	12
schools	3	391	0	0	0	391	schools	3	61	0	0	0	61
Total	570	35	29	97	731		Total	85	11	7	131	234	

This table is based on retrospective schooling histories collected in 2009/10. Years spent in kindergarten classes in private school are added to years of private schooling.

data: about 16% of the younger cohort children have changed school at least once by the age of 8 years in 2009; in urban areas, about 13.5% of this move is across sectors while in rural areas about 58% of the move is across sectors, the bulk of which is a movement from government to private schools.

Table 9 presents the years of schooling in the private and government sectors in the three academic years 2007/8, 2008/9 and 2009/10 by children in the sample.³⁹ There is both migration across sectors and different doses

³⁹In the 8-year old sample, I adopt a slightly different procedure in counting years in school in government and the private school sectors: whereas I treat enrolment in pre-primary grades ('kindergarten' classes) in private schools as being in school, I do not count enrolment in public preschools as such. This is due to the way

of treatment depending on any years that the child may have been out of school in this period. My approach in dealing with these issues is limited by what is feasible by the data. As a first pass, I estimated a variant of the VA model presented in Section 4 where, instead of private school attendance in 2009/10, I include the years of schooling in private schools and government schools between 2007/8 and 2009/10 school years. Results are presented in Table 10. As may be noted, the qualitative results are not different from the results presented in Table 4.⁴⁰

An alternative approach is to estimate the VAMs from Section 4 on only the subsample that has stayed in school for the entire period and not migrated across sectors. This subsample is similar to the sample that would have been obtained if, as in several other surveys, children were first surveyed in schools and then followed in later rounds only conditional on

preschools are arranged in the two sectors: whereas kindergarten classes are frequently integrated in private primary schools and focus on learning, government preschools (anganwadis) are more akin to daycare centres without a clear focus on learning and in typically other surroundings than the school environment. For greater detail on the preschool to school transitions in the Young Lives data, please see Singh (2014) and Woodhead, Ames, Vennam, Abebe, and Streuli (2009).

⁴⁰The specification implicitly imposes a common coefficient on each year of private or government schooling regardless of the timing of the treatment: thus, for example, attending a private school in 2007/8 or in 2009/10 are taken to have the same effect on test scores in 2010. This is a strong assumption that will be violated by any imperfect persistence of test scores which as discussed previously is very likely. It is hard to think however, why such decay will differentially affect private school students compared to government school students and thus our conclusions about the relative contribution of private vs. government schools should not be affected. Therefore I have chosen this specification both due to ease of presentation and interpretation and because sample sizes are adequate to make the relevant comparisons.

The most general specification possible is to include a dummy variable for being in private or government school in each year from 2007/8 to 2009/10 (inclusive) with out of school children as the base. However, the variation in schooling status from year-to-year (in terms of migrants either in/out of schooling or across schooling sectors) is not adequate to reliably estimate, for example, the effect of a year in private school in 2007/8 vs. 2008/9.

Table 10: Estimating differential effectiveness of a year of schooling in private vs. government schools

(a) 8-year old

VARIABLES	Rural				Urban			
	Without time use		With time use		Without time use		With time use	
	Math	PPVT	Math	PPVT	Math	PPVT	Math	PPVT
Years in Public schools after 2006/7	0.62*** (0.072)	0.23*** (0.035)	0.60*** (0.073)	0.22*** (0.033)	0.65** (0.21)	0.30** (0.10)	0.63** (0.21)	0.26* (0.12)
Years in Private schools after 2006/7	0.59*** (0.079)	0.30*** (0.030)	0.54*** (0.076)	0.28*** (0.031)	0.62** (0.21)	0.29* (0.12)	0.59* (0.22)	0.26 (0.12)
Observations	1,431	1,270	1,431	1,270	458	273	458	273
R-squared	0.405	0.326	0.429	0.337	0.357	0.219	0.374	0.258
F statistic: Equal effect	1.84	9.80	8.36	5.86	0.45	0.029	0.69	0.0097
Prob >F	0.20	0.0074	0.012	0.030	0.54	0.87	0.45	0.93

(b) 15-year olds

VARIABLES	Rural						Urban					
	Without time use			With time use			Without time use			With time use		
	Math	PPVT	Telugu	Math	PPVT	Telugu	Math	PPVT	Telugu	Math	PPVT	Telugu
Years in Public schools after 2006/7	0.38*** (0.043)	0.18*** (0.021)	0.42*** (0.028)	0.41*** (0.070)	0.16*** (0.050)	0.42*** (0.084)	0.37*** (0.052)	0.20* (0.081)	0.37 (0.18)	0.25* (0.11)	0.13 (0.13)	0.16 (0.14)
Years in Private schools after 2006/7	0.46*** (0.048)	0.24*** (0.038)	0.47*** (0.035)	0.49*** (0.077)	0.21*** (0.046)	0.48*** (0.086)	0.39*** (0.065)	0.18* (0.075)	0.35** (0.12)	0.29** (0.084)	0.12 (0.14)	0.11 (0.12)
Observations	729	715	706	729	715	706	233	162	136	233	162	136
R-squared	0.571	0.629	0.539	0.579	0.635	0.553	0.515	0.501	0.382	0.533	0.512	0.426
F statistic: Equal effect	15.2	3.68	4.92	12.5	2.95	6.59	0.45	0.33	0.050	0.60	0.25	0.46
Prob >F	0.0016	0.076	0.044	0.0033	0.11	0.022	0.54	0.60	0.83	0.48	0.64	0.53

This table is based on retrospective schooling histories collected in 2009/10. Years spent in kindergarten classes in private school are added to years of private schooling.

still being enrolled there.⁴¹ Coefficients in rural areas are higher (although not significantly so) when the sample is restricted only to those who have

⁴¹ Although it should be noted that sample sizes reduce importantly, especially for the younger cohort, due to the exclusion of children who have been in school only for a part of the three academic years up until 2009/10.

stayed within the same sector.⁴² Results for rural areas are presented in Appendix table A.4.

The checks above seem somewhat reassuring that, whereas switching does complicate the estimation of private school effects, it does not alter our basic finding that private school effects along these dimensions are modest, if at all significant.

5.4 Lag structure

Finally, I estimate specifications which test for the possibility of a different lag structure in the VAMs: specifically, I estimated the main regression specifications by including also lagged measures from time $t - 2$ instead of just a single period lag and, separately, including a third-order polynomial of the lag (as in Deming et. al. forthcoming, Chetty et. al. forthcoming and Kane and Staiger 2008) instead of the lag only in levels; coefficients on the private school premium seem stable and unchanged. These results are available on request but have not been included in the paper.

⁴²Migration across school sectors may be thought of as measurement error leading, if migration is ignorable, to attenuation bias which would cause coefficients to be biased towards zero. To the extent that there is selection into migration based on characteristics that also contribute to learning, it is reasonable to expect any such unobservables driving migration into private schools to be positively correlated with learning outcomes; for example James and Woodhead (2014) document through numerous qualitative interviews that migration is related to parental aspirations, incomes, voice and information. In this scenario, the difference in coefficients between the full sample and the restricted subsample in Table 8 overstates the effect of measurement error in the treatment variable.

The pattern noted above, of coefficients increasing only very modestly when restricting the sample to those who were in school for all three years compared to the full sample including all children, is consistent with low persistence; with a persistence of, say, 0.2 from year-to-year only 4% of the private school effect from 2007/8 will remain till the end of the 2009/10 school year, which would be consistent with the pattern of no distinguishable change in coefficients.

6 Do private schools have an impact on psychosocial skills?

Recent literature within economics (Cunha and Heckman, 2008; Cunha, Heckman, and Schennach, 2010) and an older literature in child psychology (e.g. Bandura, 1982, 1993) has stressed that psychosocial skills (sometimes called ‘non-cognitive’ skills) have a direct impact on the production of cognitive skills and on later life outcomes. As such, it seems useful to ask whether private schools have additional effects on these skills, apart from their effect on test scores. To my knowledge, this role has not been evaluated previously but may be important to assess in any evaluation of the total effect of private schools.

In the 2011 data collection, conducted in schools of a randomly selected subset of the younger cohort children, the survey collected a battery of attitudinal items relating to their experience of schooling, their belief in their own abilities, and the extent to which they believe their academic outcomes are within their control. Of the items administered, I select statements related to two dimensions which are clearly identifiable as related to psychosocial skills given an older literature within child psychology. The first dimension is agency/locus of control which is a measure of the extent to which individuals feel in control of their future outcomes (see, e.g., Findley and Cooper, 1983). The second dimension is self-efficacy which measures the extent to which individuals believe in their own abilities in particular domains (see Bandura 1993).⁴³

⁴³These concepts are closely linked but are distinct from each other. See, in particular, the discussion in Ajzen (2002) on the difference between ‘controllability’ and

Table 11: Statements and aggregate indexes of locus of control and self-efficacy

	Government	Private	Difference
<i>Panel A. Individual statements</i>			
Agency/Locus of control			
I can do well in school if I work hard	0.76	0.82	-0.06*
I cannot do well in school, even if I try hard (-)	0.68	0.71	-0.04
Making an extra effort rarely leads to success (-)	0.75	0.71	0.04
Going to school is of no use to me (-)	0.71	0.74	-0.03
Self-efficacy			
I am really good at learning English	0.71	0.80	-0.10**
Doing maths is very difficult for me (-)	0.52	0.70	-0.17***
I am proud of my achievements in school	0.66	0.67	-0.01
I can do my classwork at school without help	0.72	0.82	-0.11**
I am really good at learning Telugu	0.56	0.64	-0.08
I am really good at learning maths	0.79	0.85	-0.05
<i>Panel B. Aggregated indexes</i>			
Agency/locus of control index	-0.08	0.01	-0.08
Self-efficacy index	-0.15	0.20	-0.35***
N	559	207	766

Negatively coded answers, indicated by (-) have been reversed in this table and in the generation of composite indexes. An affirmative response indicates a positive outcome for all statements as displayed in table above. Statements are aggregated into composite indexes by averaging across non-missing responses and normalizing to have a mean of zero and a standard deviation of 1. The table only includes children in rural areas covered in the 2011 school based data collection. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The self-efficacy scale is defined using six statements and the agency/locus-of-control scale using four statements. Raw responses on the statements are provided in Panel A of Table 11, reducing the statements to whether or not a child agreed with a particular statement. As can be seen, on all statements private school students report more positive self-efficacy and in Maddux and Gosselin (2002) about differences between self-efficacy and numerous related concepts.

assessments although this difference is not always statistically significant. Panel B aggregates the statements into two indexes - as can be seen, whereas differences in the agency/locus-of-control index are not statistically significant, differences in self-efficacy are.

In order to assess if these differences may plausibly be contributed by private schools, I estimate the following specification:

$$P_{i,2011} = \alpha + \beta_1 \cdot Private_{i,2011} + \beta_2 \cdot site_i + \beta_3 \cdot X_{i,2010} + \beta_4 \cdot Y_{i,2007} + \epsilon_{it} \quad (8)$$

where $P_{i,2011}$ is the relevant psychosocial outcome variable; $site_i$ is a vector of site dummies; X is a vector of background characteristics which includes dummy variables for being male and for different caste groups, maternal and paternal years of education and a wealth index for the household; $Y_{i,2007}$ includes the mathematics and PPVT test scores from 2007; and $\epsilon_{i,t}$ is an idiosyncratic error term. As previously, I enter the right-hand-side variables sequentially. I have chosen to include test scores from 2007 (instead of more recent measures from 2010) because they predate school entry for most of the sample with the rest being in their first year of formal schooling.⁴⁴

⁴⁴Unlike test scores, in which we expect progression annually, it is not obvious that psychosocial outcomes may also be expected to improve yearly even in good schools. If, say, some schools instill better motivation in children at the start of schooling and this is contained in their subsequent test performance (as is likely), then controlling for test scores after school entry will produce a downward biased estimate of the productivity of schools regarding these skills.

Identification in this case rests on an assumption of conditional exogeneity, when controlling for test scores.

Table 12: The impact of private schooling on psychosocial skills - Rural areas, 9 years

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Agency/Locus-of-control index			Self-efficacy index		
Private School	0.16 (0.11)	0.059 (0.15)	0.030 (0.16)	0.30*** (0.070)	0.099 (0.092)	0.049 (0.099)
Mother's education level		0.0062 (0.0062)	0.0047 (0.0057)		-0.0016 (0.011)	-0.0060 (0.0096)
Father's education level		0.0076 (0.0063)	0.0063 (0.0065)		0.0068 (0.0068)	0.0055 (0.0068)
Male		0.019 (0.088)	0.028 (0.087)		0.14* (0.070)	0.14* (0.071)
Eldest child		0.12* (0.062)	0.11 (0.063)		0.079 (0.10)	0.050 (0.11)
Scheduled Castes		0.014 (0.19)	0.015 (0.20)		-0.046 (0.12)	-0.039 (0.15)
Scheduled Tribes		0.11 (0.15)	0.11 (0.16)		-0.48** (0.21)	-0.55** (0.23)
Other Backward Classes		-0.079 (0.13)	-0.071 (0.13)		-0.18** (0.080)	-0.15 (0.100)
Wealth index		0.33 (0.25)	0.32 (0.27)		0.77 (0.46)	0.60 (0.48)
Household size		0.00091 (0.018)	0.0026 (0.019)		0.033** (0.012)	0.037** (0.013)
CDA Quantitative score (2007)			0.072 (0.043)			0.12* (0.055)
PPVT score (2007)			0.012 (0.039)			0.11* (0.054)
Constant	-0.10*** (0.029)	-0.32* (0.16)	-0.30* (0.16)	-0.13*** (0.019)	-0.59*** (0.18)	-0.47** (0.20)
Observations	766	766	752	766	766	752
R-squared	0.146	0.160	0.162	0.098	0.140	0.169

Standard errors, in parentheses, are clustered at sub-district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Regression includes sub-district fixed effects. Results are from rural sample of the 2011 data collection, collected at the school.

Results from the estimation for both indexes are presented in Table 12. As is evident, the cross-sectional difference in the indexes drops with the inclusion of background controls and lagged achievement and is statistically insignificant for both indexes. I do not find any evidence for a differential effect of private schools on these measures.

7 Conclusion

In this paper, I investigated the extent of test score gaps between students of private and government schools across several cognitive domains for children aged 8 years, 9 years and 15 years in rural and urban areas; I have tried to isolate the extent to which any gaps might be causal effects of private schools; and I have attempted to understand the sources of learning achievement at the school level.

Raw differences in test scores between children in private and government schools are invariably substantial, statistically significant, and favour private school students. However, much of this variation seems to be a reflection of greater home investment and socio-economic background. Upon controlling for a wide ranging set of controls and prior achievement, for younger children I find evidence of substantially better performance only in English and a somewhat smaller effect on receptive vocabulary. For older children, I do find significant impacts of going to private schools on their scores in mathematics and Telugu; while these differences are consistently significant, they are relatively modest at about 0.2 SD and only between 20–40 per cent of the average within-community difference in test scores. In urban areas, I find no evidence of a significant private school effect. I also do not find significant evidence of a causal private school effect (relative to public schools) on psychosocial skills.

The results have several implications of interest for policy-makers. Combined with previous work highlighting that the average cost per child in rural private schools is a fraction of the average cost in the state schools, and that private schools dedicate less instructional time to Telugu and

mathematics, they suggest strongly that private schools are considerably more productive than government schools. However, they also imply that the spread of private schools is unlikely to raise average achievement levels as measured by mathematics skills or functional literacy significantly; with the exception of English, I do not find any large and consistently positive ‘private school effect’. To the extent that the first-order concern for education policy in India remains the abysmally low levels of achievement in general, rather than the inefficiencies in the delivery of education services, the spread of private schooling by itself is clearly not an adequate solution.⁴⁵

The large and significant private school premium in English, provided without any trade-off in other subjects in the case of Telugu-medium private schools and only a modest trade-off in English-medium schools, could lead to a possibly large wage premium for private school students in the future. Combined with the selectivity on socio-economic background in the private sector, this premium provides possible grounds for concern that private schools hinder social mobility and facilitate the intergenerational persistence of socio-economic status.

This study also attempts to contribute methodologically to the literature. The close correspondence of the results with experimental estimates on a comparable sample from Muralidharan and Sundararaman (2013a), as

⁴⁵The dichotomy between the two objectives - raising overall achievement and reducing the per-unit cost of achievement production - is partly artificial: it is, of course, possible (even desirable) for policy-makers to be concerned about both of these objectives. However, it needs to be acknowledged that the two objectives are distinct. Supportive evidence that the rapid rise in private schooling will only partially, if at all, address the problem of low achievement in Indian schools is also provided by the two major trends documented in the 2012 ASER Report (Pratham, 2013): while private schools have rapidly increased in their share of enrolment in rural areas between 2005 and 2012, achievement levels across states have either stayed flat (for example, in Andhra Pradesh) or in many states even declined.

well as other robustness checks reported in the paper, provide evidence of the robustness of value-added models as a mechanism for investigating causal differences in school effectiveness, which echoes a recent literature from the US and Pakistan. This is important since experiments may not be uniformly feasible across contexts and convincing natural experiments may not be available in many situations where evaluating a relevant policy question remains important. In the Indian context, it is additionally important to evaluate the robustness of VAMs since it is plausible that, if public schools do move towards a system of targeting learning outcomes instead of merely inputs (as advocated by, for example, Muralidharan 2013), then panel data on child learning outcomes would be available across the Indian public educational system. In such a scenario, it is likely that the use of VAMs will proliferate - it is worthwhile to know the extent to which such analytical exercises will be reliable and useful.

References

- AJZEN, I. (2002): “Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior¹,” *Journal of Applied Social Psychology*, 32(4), 665–683.
- AKRESH, R., E. BAGBY, D. DE WALQUE, AND H. KAZIANGA (2012): “Child ability and household human capital investment decisions in Burkina Faso,” *Economic Development and Cultural Change*, 61(1), 157–186.
- ALDERMAN, H., J. R. BEHRMAN, D. R. ROSS, AND R. SABOT (1996a): “Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy,” *Journal of Human Resources*, 31(1), 229–254.
- (1996b): “The returns to endogenous human capital in Pakistan’s rural wage labour market,” *Oxford Bulletin of Economics and statistics*, 58(1), 29–55.
- ALDERMAN, H., P. F. ORAZEM, AND E. M. PATERNO (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.
- ANDRABI, T., J. DAS, A. I. KHWAJA, AND T. ZAJONC (2011): “Do value-added estimates add value? Accounting for learning dynamics,” *American Economic Journal: Applied Economics*, 3(3), 29–54.
- ANGRIST, J. D., P. A. PATHAK, AND C. R. WALTERS (2013): “Explaining Charter School Effectiveness,” *American Economic Journal: Applied Economics*, 5(4), 1–27.

- ATHERTON, P., AND G. KINGDON (2010): “The relative effectiveness and costs of contract and regular teachers in India,” Working Paper 15, Centre for the Study of African Economies (CSAE) University of Oxford.
- AZAM, M., A. CHIN, AND N. PRAKASH (2013): “The returns to English-language skills in India,” *Economic Development and Cultural Change*, 61(2), 335–367.
- BANDURA, A. (1982): “Self-efficacy mechanism in human agency,” *American Psychologist*, 37(2), 122–147.
- (1993): “Perceived self-efficacy in cognitive development and functioning,” *Educational Psychologist*, 28(2), 117–148.
- BANERJEE, A. V., S. COLE, E. DUFLO, AND L. LINDEN (2007): “Remedying Education: Evidence from Two Randomized Experiments in India,” *The Quarterly Journal of Economics*, 122(3), 1235–1264.
- BOISSIERE, M., J. B. KNIGHT, AND R. H. SABOT (1985): “Earnings, schooling, ability, and cognitive skills,” *The American Economic Review*, 75(5), 1016–1030.
- BOLD, T., M. KIMENYI, G. MWABU, AND J. SANDEFUR (2011): “The High Return to Private Schooling in a Low-Income Country,” Working Paper 279, Center for Global Development.
- CHAUDHURY, N., J. HAMMER, M. KREMER, K. MURALIDHARAN, AND F. H. ROGERS (2006): “Missing in action: teacher and health worker absence in developing countries,” *The Journal of Economic Perspectives*, 20(1), 91–116.

- CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014): “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” *American Economic Review*, 104(9), 2593–2632.
- CHUDGAR, A., AND E. QUIN (2012): “Relationship between private schooling and achievement: Results from rural and urban India,” *Economics of Education Review*, 31(4), 376 – 390.
- CUETO, S., AND J. LEON (2013): “Psychometric characteristics of cognitive development and achievement instruments in Round 3 of Young Lives,” *Young Lives Technical Note*, 25.
- CUETO, S., J. LEON, G. GUERRERO, AND I. MUÑOZ (2009): “Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives,” *Young Lives Technical Note*, 15.
- CUNHA, F., AND J. J. HECKMAN (2008): “Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation,” *Journal of Human Resources*, 43(4), 738–782.
- CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): “Estimating the technology of cognitive and noncognitive skill formation,” *Econometrica*, 78(3), 883–931.
- DAS, J., S. DERCON, J. HABYARIMANA, P. KRISHNAN, K. MURALIDHARAN, AND V. SUNDARARAMAN (2013): “School Inputs, Household Substitution, and Test Scores,” *American Economic Journal: Applied Economics*, 5(2), 29–57.

- DAS, J., AND T. ZAJONC (2010): “India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement,” *Journal of Development Economics*, 92(2), 175–187.
- DEMING, D. J., J. S. HASTINGS, T. J. KANE, AND D. O. STAIGER (2014): “School Choice, School Quality, and Postsecondary Attainment,” *American Economic Review*, 104(3), 991–1013.
- DESAI, S., A. DUBEY, R. VANNEMAN, AND R. BANERJI (2008): “Private Schooling in India: A New Educational Landscape,” in *India Policy Forum*, vol. 5, pp. 1–58. Global Economy and Development Program, The Brookings Institution.
- DUFLO, E., P. DUPAS, AND M. KREMER (2012): “School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools,” Working Paper 17939, National Bureau of Economic Research.
- FINDLEY, M. J., AND H. M. COOPER (1983): “Locus of control and academic achievement: A literature review,” *Journal of Personality and Social Psychology*, 44(2), 419.
- FIORINI, M., AND M. P. KEANE (2014): “How the Allocation of Children’s Time Affects Cognitive and Non-Cognitive Development,” *Journal of Labor Economics*, 32(4), pp. 787–836.
- FRENCH, R. J., AND G. G. KINGDON (2010): “The relative effectiveness of private and government schools in Rural India: Evidence from

- ASER data,” Working Paper 10-03, Department of Quantitative Social Science-Institute of Education, University of London.
- GLEWWE, P. (1996): “The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment,” *Journal of Development Economics*, 51(2), 267–290.
- GLEWWE, P., E. A. HANUSHEK, S. HUMPAGE, AND R. RAVINA (2013): “School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010,” *Education Policy in Developing Countries*, 4(14,972), 13.
- GLEWWE, P., AND M. KREMER (2006): “Schools, teachers, and education outcomes in developing countries,” *Handbook of the Economics of Education*, 2, 945–1017.
- GOLDHABER, D., AND D. CHAPLIN (2012): “Assessing the ‘Rothstein Falsification Test’: Does it Really Show Teacher Value-Added Models Are Biased?,” Working Paper 2011-5, Center for Education Data and Research, University of Washington.
- GUARINO, C., M. D. RECKASE, AND J. M. WOOLDRIDGE (forthcoming): “Can value-added measures of teacher performance be trusted?,” *Education Finance and Policy*.
- HEADY, C. (2003): “The effect of child labor on learning achievement,” *World Development*, 31(2), 385–398.
- HSIEH, C.-T., AND M. URQUIOLA (2006): “The effects of generalized school choice on achievement and stratification: Evidence from Chile’s voucher program,” *Journal of Public Economics*, 90(8), 1477–1503.

- JACOB, B. A., L. LEFGREN, AND D. P. SIMS (2010): “The Persistence of Teacher-Induced Learning,” *Journal of Human Resources*, 45(4), 915–943.
- JAMES, Z., AND M. WOODHEAD (2014): “Choosing and changing schools in India across private and government sectors: Young Lives evidence from Andhra Pradesh,” *Oxford Review of Education*, 40(1), 73–90.
- JIMENEZ, E., M. E. LOCKHEED, AND V. PAQUEO (1991): “The relative efficiency of private and public schools in developing countries,” *The World Bank Research Observer*, 6(2), 205–218.
- KANE, T. J., D. F. MCCAFFREY, T. MILLER, AND D. O. STAIGER (2013): “Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project.,” Discussion paper, Bill & Melinda Gates Foundation.
- KANE, T. J., AND D. O. STAIGER (2008): “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation,” Working Paper 14607.
- KINGDON, G. (2010): “The impact of the sixth pay commission on teacher salaries: assessing equity and efficiency effects,” RECOUP working paper no. 29, University of Cambridge.
- KINGDON, G., AND F. TEAL (2010): “Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach,” *Journal of Development Economics*, 91(2), 278–288.

- KINGDON, G. G. (1996): “The Quality and Efficiency of Private and Public Education: A Case-Study of Urban India,” *Oxford Bulletin of Economics and Statistics*, 58(1), 57–82.
- (2007): “The progress of school education in India,” *Oxford Review of Economic Policy*, 23(2), 168–195.
- KINGDON, G. G., AND F. TEAL (2007): “Does performance related pay for teachers improve student performance? Some evidence from India,” *Economics of Education Review*, 26(4), 473–486.
- KINSLER, J. (2012): “Assessing Rothstein’s critique of teacher value-added models,” *Quantitative Economics*, 3(2), 333–362.
- KREMER, M., C. BRANNEN, AND R. GLENNERSTER (2013): “The Challenge of Education and Learning in the Developing World,” *Science*, 340(6130), 297–300.
- KUMRA, N. (2008): “An Assessment of the Young Lives sampling approach in Andhra Pradesh, India,” *Young Lives Technical Note 2*.
- MADDUX, J. E., AND J. T. GOSSELIN (2002): “Self-efficacy,” *Handbook of positive psychology*, pp. 277–287.
- MCEWAN, P. J. (2013): “Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments,” *unpublished*.
- MUNSHI, K., AND M. R. ROSENZWEIG (2006): “Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy,” *The American Economic Review*, 96(4), 1225–1252.

——— (2013): “Priorities for Primary Education Policy in India’s 12th Five-year Plan,” in *NCAER-Brookings India Policy Forum*, vol. 2013.

MURALIDHARAN, K., AND M. KREMER (2008): “Public and Private Schools in Rural India,” in *School choice international: Exploring public-private partnerships*, ed. by R. Chakrabarti, and P. E. Peterson. The MIT Press.

——— (2013a): “The Aggregate Effect of School Choice - Evidence from a two-stage experiment in India,” Working Paper 19440, National Bureau of Economic Research.

——— (2013b): “Contract Teachers: Experimental Evidence from India,” Working Paper 19440, National Bureau of Economic Research.

PAL, S. (2010): “Public infrastructure, location of private schools and primary school attainment in an emerging economy,” *Economics of Education Review*, 29(5), 783–794.

PRATHAM (2013): *Annual Status of Education Report 2012*. Pratham, New Delhi.

ROTHSTEIN, J. (2010): “Teacher quality in educational production: Tracking, decay, and student achievement,” *The Quarterly Journal of Economics*, 125(1), 175–214.

SINGH, A. (2014): “Test score gaps between private and government sector students at school entry age in India,” *Oxford Review of Education*, 40(1), 30–49.

- TODD, P. E., AND K. I. WOLPIN (2003): “On the Specification and Estimation of the Production Function for Cognitive Achievement,” *The Economic Journal*, 113(485), F3–F33.
- (2007): “The production of cognitive achievement in children: Home, school, and racial test score gaps,” *Journal of Human capital*, 1(1), 91–136.
- VAN DER LINDEN, W. J., AND R. K. HAMBLETON (1997): “Item response theory: Brief history, common models, and extensions,” in *Handbook of Modern Item Response Theory*, ed. by W. J. Van der Linden, and R. K. Hambleton, pp. 1–28. Springer Verlag.
- WOODHEAD, M., P. AMES, U. VENNAM, W. ABEBE, AND N. STREULI (2009): “Equity and quality? Challenges for early childhood and primary education in Ethiopia, India and Peru,” Working Paper 55, Bernard van Leer Foundation, The Hague.

Appendices to Chapter 2

Table A.1: Private school effect regressions: 8-years, rural areas

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Mathematics				PPVT		
Private school in 2009/10	0.49*** (0.11)	0.34*** (0.090)	0.14* (0.073)	0.076 (0.073)	0.45*** (0.081)	0.35*** (0.062)	0.27*** (0.067)	0.17** (0.066)
Mother's education level			0.0059 (0.0046)	0.0050 (0.0042)			0.012** (0.0054)	0.0052 (0.0044)
Male			0.033 (0.051)	0.054 (0.056)			0.14*** (0.046)	0.13*** (0.044)
Eldest child			0.019 (0.046)	0.0049 (0.052)			0.055 (0.047)	0.068 (0.046)
Scheduled Caste			0.029 (0.092)	0.033 (0.084)			0.062 (0.079)	0.038 (0.063)
Scheduled Tribes			-0.28*** (0.083)	-0.26*** (0.077)			-0.24* (0.12)	-0.20* (0.10)
Other Backward Classes			-0.028 (0.072)	-0.017 (0.067)			0.0076 (0.12)	-0.017 (0.10)
Wealth index			1.13*** (0.18)	0.99*** (0.17)			0.85*** (0.12)	0.56*** (0.14)
<i>Time use (hours per day)</i>								
Caring for others				0.061 (0.067)				-0.060 (0.044)
Domestic tasks				0.062 (0.049)				0.024 (0.034)
At school				0.12*** (0.029)				0.016 (0.035)
Studying outside of school time				0.15*** (0.029)				0.100*** (0.029)
Tasks on family farm or other business				-0.28 (0.22)				0.0023 (0.094)
Paid work outside of household				-0.067 (0.073)				-0.058 (0.17)
CDA score (lagged)		0.32*** (0.033)	0.29*** (0.029)	0.28*** (0.029)		0.27*** (0.040)		
PPVT score (lagged)								0.25*** (0.046)
Constant	-0.18*** (0.031)	-0.11*** (0.023)	-0.57*** (0.12)	-1.73*** (0.25)	0.55*** (0.018)	0.61*** (0.011)	0.098 (0.13)	0.043 (0.32)
Observations	1,438	1,438	1,438	1,438	1,305	1,305	1,305	1,275
R-squared	0.220	0.298	0.330	0.357	0.213	0.284	0.253	0.319

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Standard errors clustered at mandal level. All regressions control for mandal fixed effects.

Table A.2: Private school effect regressions: 9-years, rural areas

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
		Mathematics				Telugu				English (IRT)			English (raw)
Private school	0.49*** (0.13)	0.20** (0.084)	0.077 (0.056)	0.031 (0.061)	0.12 (0.10)	-0.089 (0.092)	-0.20* (0.10)	-0.25** (0.11)	1.02*** (0.13)	0.83*** (0.12)	0.71*** (0.15)	0.67*** (0.14)	0.46*** (0.090)
Male			0.051 (0.070)	0.052 (0.082)			-0.099 (0.085)	-0.11 (0.081)			-0.095 (0.075)	-0.092 (0.067)	0.091 (0.078)
Eldest child			0.081 (0.066)	0.12 (0.070)			0.15* (0.072)	0.18** (0.079)			0.067 (0.087)	0.081 (0.094)	-0.053 (0.084)
Scheduled Castes			-0.21 (0.14)	-0.21 (0.14)			-0.12 (0.15)	-0.12 (0.15)			-0.12 (0.10)	-0.12 (0.099)	-0.10 (0.10)
Scheduled Tribes			-0.18 (0.25)	-0.15 (0.26)			-0.14 (0.16)	-0.11 (0.16)			-0.35* (0.18)	-0.34* (0.17)	-0.44*** (0.12)
Other Backward Classes			-0.29* (0.15)	-0.28* (0.15)			-0.26** (0.11)	-0.25** (0.11)			-0.16** (0.055)	-0.15*** (0.048)	0.037 (0.077)
Mother's education level			-0.0012 (0.0048)	-0.0010 (0.0051)			0.0039 (0.0052)	0.0041 (0.0055)			0.0017 (0.0044)	0.0022 (0.0045)	0.0029 (0.0062)
Father's education level			0.0076 (0.0051)	0.0072 (0.0049)			0.0044 (0.0047)	0.0048 (0.0044)			0.0021 (0.0057)	0.0026 (0.0054)	0.0026 (0.0081)
Home support index – normalised			0.045 (0.029)	0.039 (0.030)			0.12*** (0.037)	0.11*** (0.035)			0.064 (0.037)	0.062 (0.037)	-0.045 (0.053)
Wealth index			0.44* (0.22)	0.41* (0.21)			0.43 (0.27)	0.39 (0.27)			0.62 (0.44)	0.58 (0.46)	0.32 (0.32)
<i>Time use (hours per day)</i>													
Caring for others				-0.17*** (0.056)				-0.031 (0.064)				0.030 (0.057)	-0.019 (0.11)
Domestic tasks				0.0074 (0.058)				-0.044 (0.033)				-0.025 (0.052)	0.064 (0.070)
Studying outside of school time				0.067* (0.033)				0.10*** (0.032)				0.12*** (0.020)	0.011 (0.044)
Tasks on family farm or other business				-0.22 (0.29)				0.16 (0.17)				-0.045 (0.26)	-0.055 (0.31)
Paid work outside of household				0.10 (0.14)				0.56** (0.23)				0.32 (0.25)	0.26** (0.12)
Mathematics score (lagged)		0.80*** (0.053)	0.77*** (0.053)	0.76*** (0.050)									
PPVT score (lagged)						0.52*** (0.079)	0.48*** (0.077)	0.47*** (0.075)		0.37*** (0.067)	0.33*** (0.066)	0.31*** (0.062)	0.24*** (0.063)
Constant	0.46*** (0.035)	0.56*** (0.043)	0.53*** (0.17)	0.44* (0.21)	-0.016 (0.027)	-0.30*** (0.053)	-0.28 (0.17)	-0.42** (0.18)	-0.43*** (0.036)	-0.65*** (0.051)	-0.72*** (0.19)	-0.91*** (0.22)	-0.50** (0.19)
Observations	769	769	769	769	768	711	711	711	729	672	672	672	712
R-squared	0.179	0.492	0.505	0.512	0.146	0.310	0.344	0.354	0.323	0.377	0.400	0.412	0.256

Robust standard errors in parentheses.

All regressions include mandal fixed effects.

Table A.3: Private school effect regressions: 15 years, rural areas

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Mathematics				PPVT				Telugu (cloze)		
Private school in 2009/10	0.44*** (0.054)	0.27*** (0.046)	0.19** (0.065)	0.19** (0.074)	0.59*** (0.10)	0.29*** (0.091)	0.17* (0.080)	0.15* (0.084)	0.42*** (0.093)	0.19** (0.074)	0.10 (0.070)	0.14* (0.075)
Not enrolled in 2009/10	-1.14*** (0.10)	-0.78*** (0.070)	-0.76*** (0.072)	-0.24 (0.16)	-0.82*** (0.089)	-0.42*** (0.047)	-0.39*** (0.047)	0.014 (0.13)	-1.13*** (0.085)	-0.85*** (0.080)	-0.85*** (0.080)	-0.19 (0.11)
Male			0.18*** (0.053)	0.22*** (0.064)			0.24** (0.082)	0.27*** (0.085)			0.031 (0.061)	0.079 (0.075)
Scheduled Caste			-0.21* (0.11)	-0.21 (0.13)			-0.26*** (0.085)	-0.26*** (0.086)			-0.20* (0.098)	-0.17* (0.099)
Scheduled Tribe			-0.083 (0.11)	-0.085 (0.11)			-0.26* (0.12)	-0.29** (0.11)			-0.063 (0.13)	-0.086 (0.13)
Other Backward Classes			-0.17** (0.073)	-0.14* (0.078)			-0.25** (0.10)	-0.26** (0.10)			-0.14* (0.067)	-0.11 (0.070)
Father's education level			0.0028 (0.0029)	0.0012 (0.0029)			0.0038 (0.0022)	0.0027 (0.0022)			0.0024 (0.0047)	0.00048 (0.0045)
Eldest child in the household			0.027 (0.053)	0.025 (0.052)			0.095** (0.037)	0.085* (0.040)			0.12 (0.071)	0.10 (0.069)
Wealth index			0.047 (0.25)	-0.048 (0.25)			0.27 (0.16)	0.22 (0.16)			0.21 (0.23)	0.076 (0.24)
Raven's test score			0.085*** (0.025)	0.085*** (0.025)			0.11** (0.045)	0.12** (0.047)			0.087*** (0.027)	0.083*** (0.023)
<i>Time use (hours per day)</i>												
Caring for others				-0.082* (0.046)				0.024 (0.020)				-0.053 (0.041)
Domestic tasks				0.074** (0.025)				0.027 (0.017)				0.044 (0.029)
Tasks on family farm or other business				-0.029 (0.021)				-0.024 (0.020)				-0.11*** (0.023)
Paid work outside of household				-0.028** (0.011)				-0.0026 (0.018)				-0.073*** (0.019)
Studying outside of school time				0.048* (0.025)				0.027 (0.018)				0.041 (0.034)
Hours per day – at school				0.040* (0.020)				0.044** (0.018)				0.011 (0.017)
Mathematics score (lagged)		0.45*** (0.047)	0.43*** (0.047)	0.42*** (0.048)								
PPVT score (lagged)						0.72*** (0.070)	0.66*** (0.067)	0.64*** (0.068)		0.54*** (0.060)	0.51*** (0.062)	0.48*** (0.060)
Constant	0.13*** (0.029)	0.11*** (0.020)	0.14 (0.11)	-0.37* (0.19)	0.17*** (0.034)	0.24*** (0.027)	0.19 (0.13)	-0.24 (0.16)	0.18*** (0.032)	0.23*** (0.026)	0.22 (0.15)	0.022 (0.20)
Observations	731	731	731	731	718	717	717	717	708	708	708	708
R-squared	0.381	0.517	0.534	0.550	0.322	0.591	0.629	0.637	0.326	0.468	0.482	0.513

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Standard errors clustered at mandal level. Mandal fixed effects included in all regressions.

Chapter 3

School Meals as a Safety Net:

An Evaluation of the Midday Meals Scheme in India

Abstract

Despite the popularity of school meals, little evidence exists on their effect on health outcomes. This study uses newly available longitudinal data from the state of Andhra Pradesh in India to estimate the impact of the introduction of a national midday meal program on anthropometric z-scores of primary school students, and investigates whether the program ameliorated the deterioration of health in young children caused by a severe drought. Correcting for self-selection into the program using a non-linearity in how age affects the probability of enrollment, we find that the program acted as a safety net for children, providing large and significant health gains for children whose families suffered from drought.

1 Introduction

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 and government-aided primary schools “within six months”.¹ By 2003, most states had started providing cooked meals in primary schools. Covering an estimated 120 million school children by 2006 Khera (2006), the Midday Meal Scheme (MDMS) now is the largest school-feeding program in the world.

The program was premised on expectations of significant gains in schooling and nutritional outcomes. It was expected that school meals would provide a powerful incentive for school enrollment and attendance. Additionally, it was envisioned that the program would reduce undernourishment among school children.²

The evidence, however, on the impact of the program on nutrition is rather thin. While there is evidence that school feeding in India and elsewhere does indeed improve the immediate nutritional intake of children (Afridi, 2010; Jacoby, 2002), there are few studies documenting the effect of school feeding programs on outcome indicators of child nutrition, and those that are available find ambiguous effects (e.g. Vermeersch and Kremer 2004).³

¹The full text of the court orders in this regard are available at www.righttofoodindia.org

²It was also expected that indirectly school-feeding would lead to improved levels of learning through various channels: by boosting attendance, by reducing ‘classroom hunger’ and thus improving concentration, and by improving the children’s overall levels of nutrition and thereby productivity. And finally, the program was hypothesized to deliver other social benefits such as the break-down of caste barriers by children of different castes eating together.

³The Vermeersch and Kremer (2004) study focused on children of preschool age in Kenya. The age of the children in their study (4-6 years) is almost identical with the age of the children in our sample, making their study useful for comparison purposes.

Furthermore, there has been no attempt, whether in the context of the Scheme in India or in the broader literature on school feeding programs, to evaluate their role in coping with large negative income shocks. As we show in our data, however, this role can be potentially very important in determining the distribution of impacts among program beneficiaries, especially since such shocks have been shown to have a large and enduring impact on future human capital outcomes in developing countries (e.g. Maccini and Yang 2009; see also the discussion in Strauss and Thomas 2007 and Ferreira and Schady 2009). This omission in the literature is also surprising given that the role of school meals as a safety net has been recognized by policy makers: in the Indian case, the Supreme Court ordered in 2004 that all children in drought-affected areas must be served a Midday Meal even during school vacations, which clearly reflects recognition of this role of school feeding. An evaluation of this role is therefore central to understanding the benefits of large-scale school feeding schemes.

This paper addresses these gaps in the existing literature. Using a recent longitudinal dataset from the state of Andhra Pradesh, we assess the impact of MDMS on the health status of children in primary schools. We aim to assess whether the program ameliorates the negative impact of weather shocks (drought) on children's health. Further, we aim to understand whether school meals only compensate for shocks which happen contemporaneously with the program or whether they mitigate the nutritional impacts of shocks experienced earlier in childhood through catch-up growth.

We analyze data from a longitudinal study of children in poverty collected by the Young Lives study in the state of Andhra Pradesh; details about

the sample, and the state of Andhra Pradesh, are presented in Section 3. The survey collected extensive information about children in two cohorts (born in 1994/95 and 2001/02 respectively) in 2002 and 2007. The school feeding program, known as the Midday Meal Scheme in India (henceforth MDMS), was introduced in the state in January 2003. In this study we focus exclusively on the younger cohort of children. These children were born during an 18-month period from January 2001 and June 2002; their average age was about 12 months in Round 1 and about 5½ years in Round 2. At the time of the second round of the survey, children in our treatment group would have received the school meals for an average duration of 9 months.

The period between the two rounds of the survey coincided with severe drought in our study areas and marked a period of acute agrarian distress in many villages. There was a very severe drought in 2002-3, following the failure of the monsoon rains between July-September 2002. This drought was nation-wide in impact and was the worst since at least 1987 and possibly much longer.⁴ Monsoon rainfall was also deficient in 2004 in India, and specifically in Andhra Pradesh, albeit to a smaller extent than in 2002.⁵ Children in our sample on average would have been around one year of age at the onset of the 2002-3 drought and about 3 years of age at the start of the 2004-5 drought.

⁴The drought of 2002 was covered in great detail in the national and state media in 2002 and 2003. See, for example, Frontline (2002), Kumar (2002), Financial Express (2003), The Hindu (2003) and Times of India (2003). It was also reported in the international press, for example, the New Scientist (Tata 2002). The Agriculture Secretary at the time was reported to have called the 2002 drought the worst in 120 years (Financial Express, 2003).

⁵See, for example, Financial Express (2004) and Mukherjee and Chakraborty (2004).

We use anthropometric z scores on two measures - weight-for-age and height-for-age - as the outcome variables to study the impact of the program on health and nutritional status. To correct for self-selection into the program, we utilize a non-linearity in enrollment induced by a change in the calendar year of birth: this strongly affects the probability of enrollment, as it is used as a ‘rule of thumb’ to determine the appropriate time for enrollment, but should not directly impact nutrition when controlling for age in the regressions. We use an indicator variable for whether the child was born in 2002 as an instrumental variable (IV) for our treatment dummy variable. Our IV is informative in the dataset, even though the treatment and comparison groups are only about two months apart in age on average, because data collection was carried out just as decisions on school enrollment were being made for the younger cohort children (who were between $4\frac{1}{2}$ and 6 years old in the second round); the non-linearity was a sufficiently strong predictor of whether children were in the treatment group at the time of the survey. Details of our identification strategy are presented in Section 4.

We find large benefits for children whose households self-report having suffered from drought between the two rounds; results from our preferred specification suggest drought exerts a substantial negative effect on both nutrition indicators but that these negative effects are entirely compensated for by the Midday Meals Scheme. We present evidence that the negative effect of droughts (and the compensatory positive effect of the Midday Meals) is only present for droughts that had happened at least 18 months before the second round of data collection, i.e. no later than 2005; this, in our opinion, is a reflection of the particular severity of the 2002-3 drought.

Since most children receiving the school meals would have enrolled in mid-2006, our results indicate that the major channel through which the compensatory effect of Midday Meals is realized is catch-up growth. We also show that the main results are robust to different identification and measurement strategies (Section 5).

The magnitude of the effects of the MDMS program is very large: for boys aged 65 months (the mean age in our sample), for example, the preferred specifications using self-reported drought suggest that drought creates a height loss of about 0.77 s.d. which roughly equals the distance between the 25th and 50th percentiles, and a weight loss of about 0.44 s.d. which equals two-thirds of the same distance, in the WHO (2007) growth charts; our results suggest that this entire gap is compensated for by the program.

While similarly large estimates of the negative impacts of shocks to the household on the anthropometric z-scores of children have previously been observed in the literature⁶, the large magnitude of the compensatory impacts of the Scheme may appear surprising. At first glance, it may also appear surprising that drought and school feeding has a significant effect not only on weight-for-age but also on height for-for-age, remediation for which is often not considered possible after the first 24-36 months of a child's life.

Catch-up growth in height-for-age has, however, been observed in large magnitudes in several countries. Our findings are broadly consistent with the medical literature on nutrition and supplementary feeding which documents that while growth deficits persist into

⁶For example, Akresh, Verwimp, and Bundervoet (2011) document a fall of 0.86 s.d. in height-for-age for girls as a result of crop failure and a fall of 1.05 s.d. in height-for-age for children who suffered from exposure to the Rwandan civil war.

early adulthood if children remain in poor conditions, there is potential for catch-up in height-for-age if circumstances improve for the better, such as through nutritional supplementation or migration when children are still young (see for example Tanner 1981; Coly, Milet, Diallo, Ndiaye, Bénéfice, Simondon, Wade, and Simondon 2006; Golden 1994). In motivating our empirical tests, we note that not only has catch-up growth in height-for-age been documented at these ages and later in childhood in different contexts such as the Philippines, Senegal and Peru, the magnitude of catch-up is frequently observed to be larger than the effect sizes we report in this paper.

This paper makes several new contributions to the literature. It is the only econometric evaluation, to our knowledge, of the effect of India's Midday Meals Scheme on the health outcomes of children; it contributes to the broader literature on school feeding as described above, including its unique focus on the impact of school feeding in coping with negative income shocks; it is, to our knowledge, one of the few papers that evaluates a plausible policy measure to facilitate catch-up growth in a context where child malnutrition is very widespread⁷; and finally, it is one of the few evaluations of the impact of school feeding that corrects for self-selection and incorporates dynamic aspects of health determination.

⁷See Haddad (2013), for instance, detailing the high rates of undernutrition in India, and the necessity for policy interventions to combat this; 40% of Indian children were stunted in 2005-6 according to the National Family Health Survey Round 3.

2 The Midday Meals Scheme in India

The Midday Meals Scheme is among the most important initiatives by the Indian government in the area of education in recent years. Under the scheme, on every school day, all primary school students in public schools are provided with a cooked meal containing no less than 300 kcal and 8-12 grams of protein.⁸

Although it was officially started in 1995, the Scheme remained unimplemented in most states until 2002. Following a Supreme Court ruling in 2001, the Scheme was implemented across the country. As such it represents, at least in terms of outreach, one of the most successful interventions by the Indian government in recent years.

Andhra Pradesh (AP) started providing Midday Meals in January, 2003 to children in all primary and upper primary public and private-aided schools. As several studies document, this Scheme was nearly universal from the very beginning. Drèze and Goyal (2003) report full implementation of the Scheme in 2003 in AP. In later years, Thorat and Lee (2005) and Pratham (2007) report that over 98% of government schools in the state were serving a Midday Meal on the day of their school surveys.⁹

Much interest was generated in the performance of the Scheme after 2001, when the issue entered the mainstream political and media discourse in

⁸From the 2006-7 school year, which is the relevant school year for our study, the minimum nutritional standards were revised upwards to at least 450 kcal and 12 grams of protein. The school meals in our study therefore, may have represented an even bigger nutritional increment to the diet of beneficiaries than estimated in previous studies.

⁹Private aided schools are run under private management but receive government funding and support, have access to government schemes like the Midday Meal Scheme, and follow the same regulations including pay and tenure regulations as government schools. In practice, their quality and functioning is often indistinguishable from public schools (Kingdon, 1996). These form a very small part of the number of schools in Andhra Pradesh, about 4% according to Mehta (2007).

India. As a result, several field studies were carried out over the next few years. Most studies of the program, with the exception of Afridi (2010, 2011) and Jayaraman and Simroth (2011), were non-econometric in nature and looked at descriptive statistics based on school records.¹⁰

Khera (2006) is the best review article of these surveys; it lists nine surveys completed during the period 2003-2005 focusing on MDMS and reviews their major findings. In general the surveys focused on the effect of the scheme on enrollment, attendance and retention as well as aspects of infrastructure change, caste discrimination and opinions of stake-holders (teachers and parents). The surveys were almost unanimous in documenting a rise in attendance rates as well as enrollment rates especially benefiting girls and, in one study, children from the scheduled castes. Afridi (2011) confirms findings on attendance using a difference-in-differences estimator, finding large benefits in school participation especially for girls. Jayaraman and Simroth (2011) document a 13% increase in enrollment in response to the Scheme, identifying the effects from timing differences in the roll-out of the Scheme across different Indian states.

Afridi (2010) is the only paper that looks at the nutritional impact of the program in India. Using a 24-hour recall of food intake in a randomised evaluation in Madhya Pradesh she found that “daily nutrient intake of program participants increases by 49% to 100% of the transfers. For as low a cost as 3 cents per child, the program reduces daily protein deficiency of participants by 100% and calorie deficiency by almost 30%.”

¹⁰See Jayaraman and Simroth (2011) for a detailed discussion of the Scheme and its implementation across different Indian states.

However, the question that we are interested in, namely that of the longer term impact of the Scheme on child health and its role as a safety net, has not been directly addressed by any previous study.

3 Data and background

The data we use in this study were collected by the Young Lives study between September and December of 2002 and between January and June of 2007 in the state of Andhra Pradesh.¹¹ Andhra Pradesh is the fourth-largest state in India by area and had a population of over 84 million in 2011. It is divided into three regions – Coastal Andhra, Rayalaseema and Telangana – with distinct regional patterns in environment, soil and livelihood patterns. Administratively the state is divided into districts, which are further sub-divided into sub-districts (mandals) which are the primary sampling units for the survey.

The surveys cover two cohorts: the first is comprised of 2011 children born between January 2001 and June 2002, and the second includes 1008 children born between January 1994 and June 1995. In the second round conducted in 2007, 1950 children of the younger cohort and 994 children of the older cohort were successfully resurveyed; attrition rates thus are low and do not

¹¹About 94% of the interviews in Round 2 were carried out between January-April 2007; children interviewed after this period were often those who had migrated outside the original Young Lives communities and thus needed to be interviewed separately from the rest of the sample.

pose a problem for the analysis.¹² In this paper, for reasons of program identification, we focus exclusively on the younger cohort.¹³

The period between 2001 and 2007 saw severe drought in several parts of the state, especially in 2002-3. In these years, districts in our sample saw a severe shortfall in rain of up to 40% below normal rainfall.¹⁴ This had a devastating impact on agricultural activity, much of which is primarily rain-fed; in 2002-3, the total food grain production in Andhra Pradesh was a quarter below the ‘normal’ production in both of the main agricultural growing seasons in the state . The drought was especially severe in Rayalaseema and Telangana regions which are particularly drought-prone.

Figure 1 presents a timeline of the major events in our data and the age of the younger cohort children at the time. The children were born in an 18-month age range between January 2001 and June 2002; failure of the monsoon rains between July and September 2002 led to a severe drought; the first round of fieldwork was carried out between September and December 2002; the 2006/7 school year began in mid-June 2006, which marks the beginning of treatment for the program beneficiaries in our sample; finally, the second round of data collection was carried out between

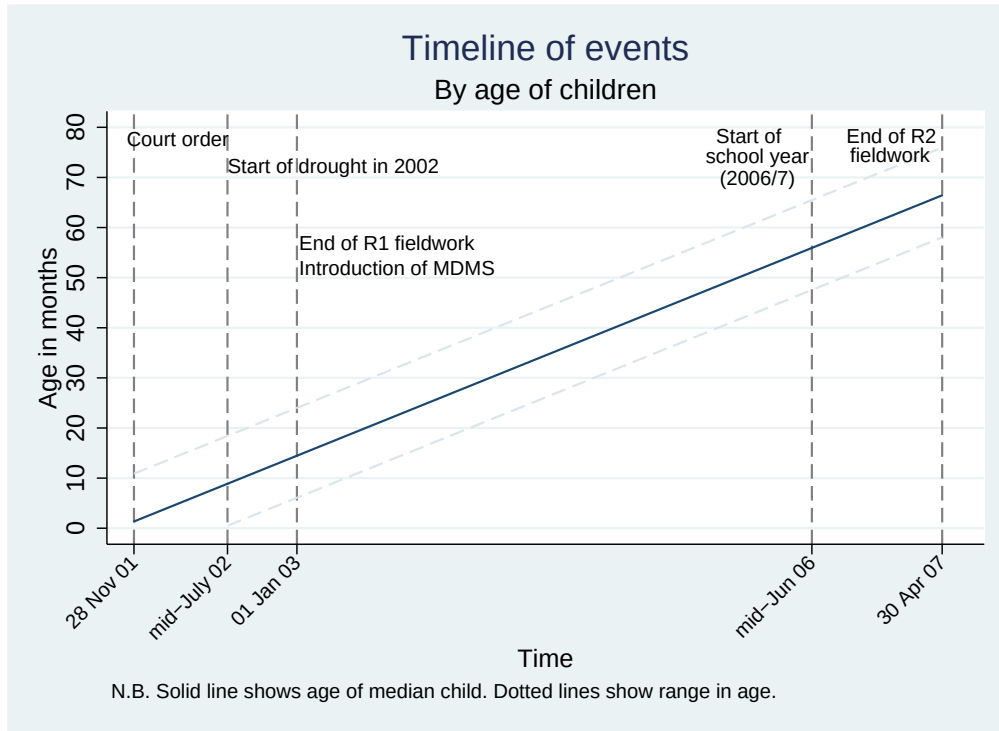
¹²For greater details about the representativeness of the Young Lives sample, as compared to the DHS sample for Andhra Pradesh, as well as details of attrition, please refer to Kumra (2008).

¹³The only feasible comparison groups for the older cohort, who were about 8 years old in Round 1 and 12 years old in Round 2, are students in private schools or not enrolled, who are likely to differ in systematic ways from students in public schools, precluding a credible identification strategy. OLS regressions, similar to those implemented for the younger cohort, did not reveal any impact of the Midday Meals on nutrition outcomes for children in the older cohort.

¹⁴These figures are based on data from the Directorate of Economics and Statistics, Government of Andhra Pradesh and were retrieved from www.indiastat.com in January 2012. Although rainfall was deficient in 2004 also, the effect on agricultural production was much smaller at around 2% below normal.

January-April 2007. The figure presents the age of the median child as well as the range of ages at these points.

Figure 1: Timeline of events



In this paper, we evaluate whether access to school-provided meals between mid-June 2006, when the children were between 4 and 5.5 years of age, and the time of the survey in 2007 had any direct impact on the anthropometric indicators of these children; further, we investigate whether participation in the program in 2006/7 compensated for negative health shocks caused by a severe drought in 2002-3 which had affected these children in infancy. The severe drought pre-dates the school feeding by over three years and thus any impacts that we find of the compensatory effects of school feeding must be operating through the channel of catch-up growth. At the time of the second round of the survey in 2007, children in our treatment group would have received the meal for an average duration of about 9 months,

with a minimum of 7 months and a maximum of one year. Children in the treatment group were aged between 48 and 65 months, with an average age of about 57 months, at the start of the 2006/7 school year.

Our goal is to investigate whether a large nutritional intervention, the Midday Meals Scheme, could compensate for early nutritional deprivations caused by drought in years prior to the intervention. The dataset has several strengths for our purposes. Firstly, it covers just the right period: the first round was in 2002 just before the program was implemented in AP. in January 2003, and the second round was in 2007, long enough for the Scheme's teething problems to have been resolved. The period spanned a period of severe drought, making the data suitable for understanding the impact of Midday Meals in cushioning the impact of drought. Secondly, the longitudinal nature of the data helps greatly in dealing with problems in estimation and identifying impact. Thirdly, the age of the younger cohort was exactly around the normal time of school enrollment; as we later discuss, this is critical to our identification of program impact. Finally, no other baseline surveys for the Indian scheme exist, to our knowledge, from which we can obtain a better estimate; this in itself makes the data important.

4 Framework and methodology

Following Senauer and Garcia (1991), Behrman and Deolalikar (1988) and Behrman and Hoddinott (2005), we conceive of child health as entering directly into the welfare function of the household, reflecting the intrinsic

value of child health to the household. Health is determined by a health production function of the form:

$$H_{it} = H(F_{it}, C_i, Z_{it}, H_{it-1}, U_{it}) \quad (1)$$

where H_{it} is the health of child i at time t , F_{it} is the child's food consumption in period t , C_i is a vector of time-invariant observable characteristics of the child, including determinants such as caste, gender, and parental education, Z_{it} is a vector of time-varying characteristics such as recent economic shocks, H_{it-1} is previous period health, and U_{it} is a vector of unobserved attributes of the child, parents, household, and community which affect the child's health status. The function allows for the possibility of interaction effects among its arguments.

Our focus here is not to estimate the structural parameters of the health production function but to evaluate the policy effect of MDMS on child malnutrition. We assume that access to the Midday Meal Scheme, captured by the binary variable $MDMS_{it}$, results in a net increase in child food intake (F_{it}), as found for example by Afridi (2010) in India and Jacoby (2002) in Philippines. Following equation (1), we estimated the following two equations:

$$H_{i2} = \alpha + \beta_1.MDMS_{i2} + \beta_2.Drought_i + \beta_3.H_{i,1} + \beta_4.C_i + \epsilon_i \quad (2)$$

$$H_{i2} = \alpha + \beta_1.MDMS_{i2} + \beta_2.Drought_i + \beta_3.MDMS_{i,2} * Drought + \beta_4.H_{i,1} + \beta_5.C_i + \epsilon_i \quad (3)$$

Here, $MDMS$ refers to the treatment dummy variable, $Drought$ refers to self-reported drought having occurred between 2002 and 2006¹⁵, $H_{i,1}$ refers to first-period nutritional z score, and C_i is a vector of other controls, including dummy variables for different castes, being male, urban location as well as household size, caregiver's education, and a household wealth index. All variables in C_i are from Round 1 (2002). Equation (2) presents the average treatment effect for the sample. Equation (3) allows for an interaction effect between $MDMS$ and drought which will allow us to explicitly investigate the role of the program as a safety net.

In considering the dynamic processes that determine health outcomes modelled in equations (2) and (3), it is important to keep in mind the timing of events and measurements described in section 3. One would probably expect that less than one year of exposure to a school meal program at the age of 4 to 5.5 years would have a larger influence on children's weight-for-age than on children's height-for-age, given that weight is usually more responsive to changes in nutrient intake in the short run.

However, predictions become more complicated when one considers the possibility for catch-up growth and how it may be related to previous nutritional deprivation. For a nutrition intervention at age 4-5.5 to reverse insults to nutrition caused by a major drought that started when the children were less than 2 years old, there must be biological evidence that catch-up growth under such circumstances is possible. Since height is a longer-term measure of health compared to weight, droughts that occurred

¹⁵We will show that our conclusions are robust to using alternative measures of drought. We will also show later, using information on the timing of the drought, that our results are not driven by contemporaneous droughts but reflect catch-up growth in those children who had suffered health deterioration from drought in early childhood.

some years ago may have a more pronounced impact on height than weight, providing more scope for compensating effects of a nutritional intervention. In contrast, recent or current droughts may have a larger impact on weight. In fact, the nutritional literature presents a number of examples of such catch-up growth. While early growth deficits often persist into adulthood, it is not established that this is a biological necessity: as Golden (1994) hypothesizes: “the available data could be interpreted to show that a period of malnutrition in the first 2–3 years irrevocably changes the child so that he is ‘locked into’ a lower growth trajectory with a lower potential for future growth. The alternative hypothesis is that full catch-up growth is possible. However, this is not observed in practice because the correct conditions are not satisfied because in most populations environment and diet do not change.” (quoted in Boersma and Wit 1997).

Furthermore, the magnitude of catch-up in the children who recover from previous health deprivations is frequently very large. Adair (1999), for example, looks at catch-up growth between 2 and 12 years using the Cebu panel study from the Philippines and finds that almost a third of the children stunted at 2 years of age experienced catch-up growth by the time they were 8.5 years old. Those who did experience catch-up growth had mean improvements in height-for-age z scores of 1.14 s.d. Similarly, Crookston, Penny, Alder, Dickerson, Merrill, Stanford, Porucznik, and Dearden (2010), analyzing Young Lives data from Peru which compares children of exactly the same age in 2002 and 2006/7 as our sample, also document catch-up for a large proportion of the children who were stunted in Round 1; for those children, in whom catch-up growth is observed, the magnitude of catch-up is an average of 1.13 s.d. Coly et. al. (2006) document a

similar magnitude of catch-up in Senegal with large changes for those stunted at preschool: using WHO norms, they find mean height-for-age z score increases of 0.21, 0.90, and 1.79 s.d. for girls, and 0.31, 0.95, and 1.44 s.d. for boys, with no stunting, mild stunting, and marked stunting, respectively.

Another relevant result from the biological literature on catch-up growth is that catch-up growth may be more difficult for children who suffer severe undernutrition in the first years of life. Adair (1999) documents in the Philippines, the likelihood of recovery from stunting is lowest for those who were stunted in the first year of their birth; similarly Crookston et al. (2010) also note that catch-up seemed to be most likely for those who had higher height-for-age scores at initial assessment. We can also test whether in our study sites in India catch up growth differs between children whose growth is stunted early in life and those whose early growth is not stunted.

Identification

Of the children in the younger cohort, who were aged between $4\frac{1}{2}$ and 6 years old in 2007, about 45% were in school by the second wave. Of these students, 79% were in public schools and the rest were in private schools (including those run by NGOs and religious charities). Most of the children who are not yet in school in the second round would join formal schooling soon; the survey therefore also asked the caregivers of children not yet in school what type of school (defined as public, private, religious etc.) their child would be likely to join and the age at which they thought the child

will be enrolled: the caregivers of over 95% of the children not yet enrolled report that they expect the child to be in school by the age of 6 years.¹⁶

In the data, only 1.47% of caregivers of the children enrolled in public schools (10 out of 682) reported that their school does not provide a midday meal, thus confirming the widespread implementation of the program indicated by previous studies.¹⁷ We therefore define the treatment group as all children currently attending public school.¹⁸ Our results are not driven by the assumption that all children in public schools receive the meal; such an assumption should indeed bias our results toward finding a weaker impact of the program. The results are unchanged if we use the availability of the meals, as reported by the caregiver, to define the treatment group. A major concern related to non-random program placement is the endogeneity of treatment (enrolling in a public school), especially via self-selection into the program. It is possible that self-selection into public schools is correlated with anticipated benefits of the program as reflected in changes in health or learning over time. Parents could have been influenced by the Midday Meals Scheme in deciding whether and at what age to enrol their children in public schools. Self-selection can take place through multiple mechanisms: attracted by the introduction of the Midday Meals parents can i) decide to send their children to a public school rather than no school at all or ii) to a public school instead of a private school, or iii) they can

¹⁶The question of when the child is expected to join the school in the future elicited responses in completed years of age and not months.

¹⁷Caregivers of another 24 students (3.52%) report not receiving the midday meal because the child does not like the food.

¹⁸The caregivers of about 98.5% of children in public schools report that the school provides the meals, indicating the ten cases of reported non-availability of food may either reflect temporary unavailability or the caregiver's lack of knowledge about the whether the child receives the meal or not.

decide to enrol their child in a public school at a younger age than they otherwise might have in order to benefit from the program.¹⁹

¹⁹The relative importance of these channels of self-selection is likely to vary across regions. The first channel is unlikely to be very important in Andhra Pradesh because nearly all children in the state go to primary school. For instance, even in the first round of data collection (before the introduction of the Midday Meal Scheme), over 97% of the children in the older cohort, then aged 8 years, were in school. We suspect the second channel also is not too important as the program is likely to be an incentive only for poorer households, and children from these households, especially in rural areas, would typically enrol in a public school anyway. It is the third channel that is most likely to be influential. That this channel is influential in at least some cases has been documented in the qualitative data collected by Young Lives – some parents do enroll their children before the official age of enrollment just so that they can benefit from the Midday Meal.

Table 1: Descriptive statistics

VARIABLES	Treatment Group	Comparison Group	Total
Male	0.511 (0.5)	0.506 (0.5)	0.509 (0.5)
Urban	0.058 (0.233)	0.103*** (0.304)	0.077 (0.267)
Drought	0.34 (0.474)	0.386* (0.487)	0.36 (0.48)
Wealth index (2002)	0.327 (0.161)	0.312 (0.16)	0.32 (0.161)
Caregiver's education	1.87 (3.02)	2.429 (5.667)	2.185 (4.708)
Household size	5.621 2.435	5.427 2.424	5.512 2.43
Scheduled Castes	0.23 (0.421)	0.211 (0.408)	0.222 (0.416)
Scheduled Tribes	0.212 (0.409)	0.164** (0.371)	0.191 (0.393)
Backward Classes	0.439 (0.497)	0.515** (0.5)	0.472 (0.499)
Other Castes	0.119 (0.325)	0.11 (0.313)	0.115 (0.32)
Telangana region	0.279 (0.449)	0.375*** (0.485)	0.321 (0.467)
Rayalaseema region	0.344 (0.475)	0.295* (0.456)	0.323 (0.468)
Coastal Andhra Pradesh	0.377 (0.485)	0.33* (0.471)	0.357 (0.479)
Height-for-age z score (2002)	-1.351 (1.461)	-1.597*** (1.53)	-1.457 (1.495)
Weight-for-age z-score (2002)	-1.621 (1.064)	-1.843*** (1.167)	-1.717 (1.115)
Height-for-age z-score(2007)	-1.645 (0.83)	-2.1*** (0.96)	-1.844 (0.917)
Weight-for-age z-score (2007)	-1.879 (0.854)	-2.181*** (0.865)	-2.011 (0.872)
Age (in years)	5.5 (0.276)	5.29*** (0.324)	5.41 (0.315)
N	695	536	1231

Note: This table presents means of variables by group; standard deviations in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In our analysis, we utilize the information on the type of school that children will join in the near future to restrict the comparison group to children who are not currently enrolled but will be enrolled in a public school in the near future; thus our preferred specification compares only children currently in public schools to children who will in the future go to public schools.²⁰ This allows us to abstract from the endogeneity of the choice between private or public schooling. In Table 1 we present summary statistics across a range of measures for the treatment group and our (restricted) comparison group of children who will join public schools in the future. There are significant differences in the mean of background variables between the treatment and the comparison groups; however these differences are frequently much smaller in magnitude and in statistical significance when using our preferred (restricted) comparison group, rather than when we compare our treatment group to all children not currently receiving the school meals.²¹

Using this sample, our treatment and comparison groups are mainly differentiated by whether they have enrolled in school; as we discussed earlier, it is plausible that this decision is endogenous and affected by the availability of the Scheme.²² To address endogeneity problems caused by self-selection in enrollment, we adopt an instrumental variable approach. The requirement for an IV in this case is that it should be able to predict enrollment in school at the time of the survey, in this sample

²⁰We do, however, also report results including all children not currently enrolled in public schools, i.e. all children not yet enrolled and those enrolled in private schools, in the comparison group.

²¹This comparison, using a comparison group of all children not currently enrolled in public schools (and thereby not receiving the meals), is presented in Appendix 1.

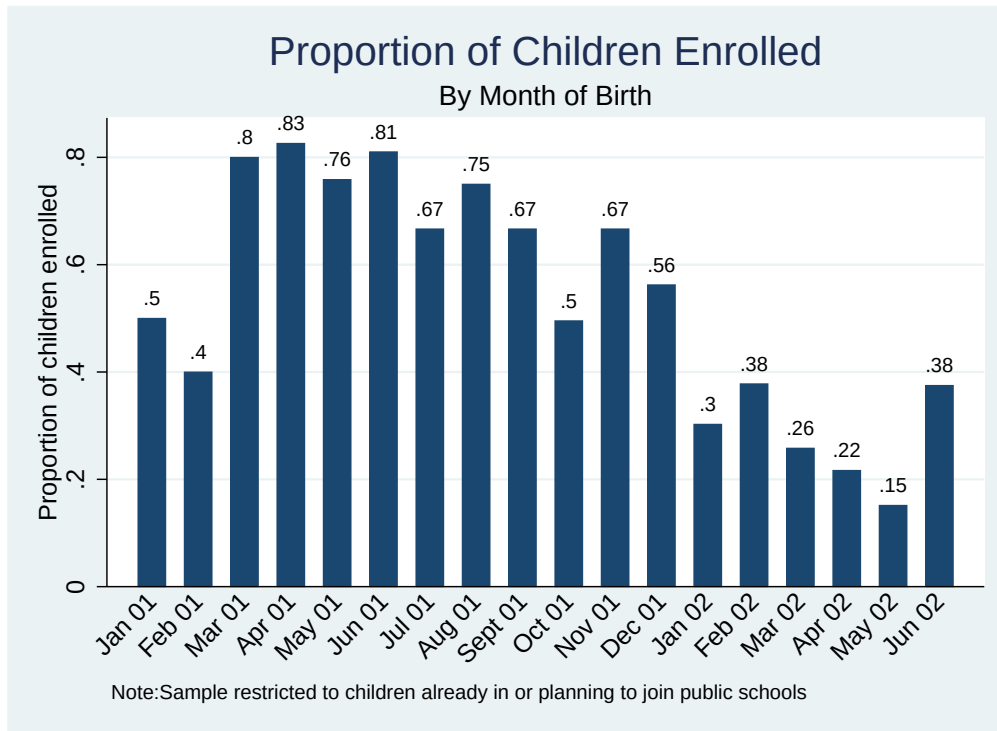
²²This concern ties in directly with the theoretical framework where we posited that unobservable factors at the household and child level might directly affect nutrition; if these unobservable factors (e.g. parental concern) directly affect the probability of enrollment as well, OLS estimates of the treatment effect would be biased. This concern prompts us to use an instrumental variable approach.

of children who are either in public schools already (treatment group) or will join public schools in the future (comparison group), but not otherwise be an independent determinant of nutrition. To this end, we exploit a non-linearity in the relationship between age and enrollment induced by a change in the calendar year of birth; this non-linearity affects the probability of enrollment at this particular point of the children's educational trajectory, but is not expected to be associated with the nutritional outcomes of children once we separately control for age.

Noting again that all children in our sample are born between January 2001 and June 2002, we create an indicator variable for being born after December 2001 and use this as an instrument that would predict enrollment but not nutrition, at the same time controlling for the linear effects of age.²³

²³Age was calculated based on the difference in days between the date of interview and the date of birth. The treated group is older on average than the comparison group in the sample, which is as we would expect; the mean difference is about two months.

Figure 2: Proportion of children enrolled by month of birth



The intuition behind our use of this variable as an IV is straight-forward. Parents and teachers often use the calendar year of a child's birth to decide when he or she should enroll in school. Although the probability of being enrolled generally increases with age, such a rule of thumb would be expected to create a nonlinearity in the relationship between time of birth and enrollment between December 2001 and January 2002. That this non-linearity is empirically important can be seen clearly in Figure 2 which plots mean enrollment rates by month of birth. The proportion of children enrolled drops nearly in half from 56% of children born in December 2001 to 30% of those born in January 2002. Although there is noise in the month-to-month variation in enrollment rates, there is a

sharply more negative relationship between birth month and enrollment rate in the months around the end of 2001.²⁴

This nonlinearity is consistent with the rule of thumb described based on calendar year of birth or could arise naturally around this time threshold due to social norms about the age of enrollment.²⁵ When other months are chosen as the threshold point for changes in enrollment probability, they are much weaker and usually lack statistical significance, suggesting that the nonlinearity in the relationship between time of birth and enrollment is specific to this time threshold.

Given the threshold nature of our instrument, our approach can be considered a regression discontinuity design in which we control for the running variable (age) with a linear term and use the discontinuity as an instrument for the treatment of interest. The linear control for age is reasonable given the limited range of birth months; when we try including higher order terms for age, the threshold variable lacks sufficient power to explain variation in enrollment rates.

²⁴That small differences in age are associated with large differences in enrollment, as in Figure 2, is a product of the specific point of their educational trajectory that the children are in i.e. at the very age that decisions about school enrollment are being taken; at any other ages outside this narrow window, we would expect to see no variation in enrollment induced by age differences of only 2 to 3 months. The lower rate of enrollment for children born between Jan-Mar 2001 seems puzzling but is explained by the fact that only 28 (~2%) children in the dataset out of 1231 children who are in or will join public schools were born in this period. Similarly, only 8 children are born in June 2002 which has a higher enrollment rate than the preceding months. Results are not sensitive to the exclusion of children born in these months. 5 years is the prevailing norm for the age of enrollment into public schooling in Andhra Pradesh. For example even in the older cohort, 70% of the children who had joined public schools by Round 1 (2002) when they were 8 years old, entered formal schooling at 5 years of age.

²⁵5 years is the prevailing norm for the age of enrollment into public schooling in Andhra Pradesh. For example even in the older cohort, 70% of the children who had joined public schools by Round 1 (2002) when they were 8 years old, entered formal schooling at 5 years of age.

Our instrumenting strategy outlined implies a first-stage equation of the form:

$$MDMS_{i2} = \mu + \pi_1 Born2002 + \pi_2 Age + \pi_3 Z + \epsilon_i \quad (4)$$

where *Born2002* is an indicator variable for being born in 2002, equalling zero if the child was born in 2001, *Age* is age at the time of the survey measured in years with daily precision, and *Z* is a vector of exogenous variables including all exogenous covariates in the second-stage equation (Equation 3), the instruments for first-period anthropometric z score (perceived size at birth and death of a household member during pregnancy) and an interaction term between *Born2002* and *Drought* variables which is used as an IV for the interaction term between *MDMS* and *Drought*.²⁶ The exclusion restriction on the IV would be violated if a change in the calendar year of birth had a non-linear impact in this age range, not only on the probability of enrollment but also on the changes in the anthropometric z scores. We do not however, have any reason to expect this to be the case: our anthropometric z scores are norm-referenced by age measured in days. Additionally, the children in the enrolled and non-enrolled groups are very close in mean age.

Furthermore, any general non-linear impacts of age should not be confined to the impacts of the scheme on drought affected children but on the entire group of beneficiaries as a whole. Our results however indicate that the entire benefit of the Scheme is concentrated on children whose households

²⁶Given the exogeneity of *Drought*, if *Born2002* is a valid IV for *MDMS*, then an interaction term of these two variables is a valid IV for the interaction term of *MDMS* and *Drought*.

reported being drought affected. One possible effect of time of birth on child health that could have affected children only in drought-affected areas is the age of exposure to the 2002-2003 drought. Children born after year-end 2001 were younger when the drought began to create hardship in the second half of 2002 (0 to 6 months old in mid-2002) and so could have been more affected by the drought than the older children in the sample, who were 6 to 18 months old in mid-2002. However, for this to be a problem, the relationship between age of exposure and health impacts of the drought must not only be nonlinear but also must be nonlinear around the specific threshold of 6 months.

A possible way to test whether the effects we find are due to nonlinear effects of age on nutrition when exposed to drought is to test for threshold effects of age on nutritional outcomes for the sample of children enrolled in or planning to enrol in private schools. Since these children did not have access to the Midday Meals Scheme, finding any nonlinear effects of age on nutritional outcomes would provide evidence against our identifying assumptions. As we describe in the next section, we do not find any evidence to this effect.

Incorporating the dynamic aspects of health determination is both desirable and essential but it exposes us to the problem of endogeneity of the lagged dependent variable. We instrument the lagged dependent variable (anthropometric score from Round 1) using the caregiver's perception of birth size and shocks during pregnancy (whether a household member had died). Birth size is related to conditions during pregnancy and is very strongly correlated with a child's health in the first 18 months of life. The instruments are appealing because they are pre-determined when the lagged

health measurement is taken and so should be less correlated with current outcomes than the more recent lagged measurement; however, we cannot rule out remaining correlation with unobserved household characteristics that affect current child health.²⁷

5 Results

For descriptive purposes, we estimate the unconditional average treatment effect on the treated (ATT) by a simple OLS regression of the change in the z-score on treatment. We ran the regression on the full sample, and also separately for children who had suffered from drought, and children who had not. Drought is the major economic shock in this region; 38.56% of households in rural areas in the restricted sample self-reported having been affected by drought between the two rounds.

Specifically we estimated equations of the form:

$$\Delta Y = \alpha + \beta_1.MDMS_i + u_i \quad (5)$$

Here, Y is the health measure and $MDMS$ the treatment binary. This merely shows the difference between the average changes in Y between the two groups. It is only intended as a first look at the data and ignores the econometric problems discussed in the previous section.

²⁷Birth weight might have been a better IV but was impracticable in this case. Birth weight was only available for about half the sample as many of the children were born at home and without medical attention. It is important to note that our results do not depend on the inclusion or instrumenting of the lagged dependent variable. The patterns around the impact of the drought, and the cushioning effect of the Midday Meals, are similar in sign and statistical significance (although with even greater magnitudes) if we redefine our estimated equation using changes in z scores as the outcome variable and omit the lagged dependent variable from the regressors.

Table 2: Unconditional treatment effect from OLS regressions on binary treatment variable

VARIABLES	Full public school sample		Restricted to drought-affected		Children not drought-affected	
	(1) Changes in weight-for-age	(2) Changes in height-for-age	(3) Changes in weight-for-age	(4) Changes in height-for-age	(5) Changes in weight-for-age	(6) Changes in height-for-age
Midday Meals	0.074 (1.32)	0.20** (2.18)	0.23*** (3.41)	0.43** (2.85)	-0.028 (-0.49)	0.034 (0.39)
Constant	-0.33*** (-5.81)	-0.48*** (-3.14)	-0.51*** (-9.19)	-0.95*** (-4.07)	-0.22*** (-3.63)	-0.19*** (-3.05)
Observations	1,215	1,193	436	422	779	771
R-squared	0.002	0.006	0.017	0.024	0.000	0.000

Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2 presents the descriptive estimates of the unconditional impact estimated by the exercise above. These initial results indicate that the treatment had a significant impact on both measures for children who had suffered from drought but not for children who did not: these preliminary estimates imply a positive benefit of 0.23 s.d. for weight-for-age and 0.43 s.d. on height-for-age z scores; there are no significant impacts on children who did not suffer from drought.

Next, we report the results of estimating (2) which captures the average impacts of treatment on the treated, not allowing for heterogeneous treatment effects. Then, we estimate (3) by adding an interaction between MDMS and an indicator variable for whether the child experienced drought in the previous four years. In both cases, we estimate the effect of MDMS both with and without the instrumentation of the MDMS dummy. As expected, results from the first-stage are strong and Born2002 significantly predicts being in the treatment group, even controlling for all covariates in Z (including age which is controlled for in all specifications).²⁸

²⁸Birth weight might have been a better IV but was impracticable in this case. Birth weight was only available for about half the sample as many of the children were born

Table 3: Estimated impact of Midday Meals on children's nutrition

VARIABLES	(1) Weight-for-age in 2006/7 OLS	(2) IV	(3) Height-for-age in 2006/7 OLS	(4) IV
Midday Meals	0.15*** (3.56)	0.60*** (3.07)	0.22*** (2.95)	0.27 (0.85)
Drought	-0.11*** (-2.88)	-0.088** (-2.05)	-0.22*** (-3.77)	-0.21*** (-3.57)
Age expressed in years	-0.090 (-1.10)	-0.34*** (-2.78)	0.38*** (2.77)	0.33 (1.58)
Weight-for-age in R1	0.61*** (7.52)	0.57*** (7.30)		
Height-for-age in R1			0.58*** (5.24)	0.56*** (5.77)
Constant	-0.54 (-1.47)	0.53 (0.93)	-3.00*** (-4.66)	-2.82*** (-2.83)
Observations	1,199	1,199	1,178	1,178
R-squared	0.378	0.341	0.220	0.237
Kleibergen-Paap F-statistic	32.8	11.9	11.2	12.9
Hansen J-statistic p value	0.83	0.40	0.39	0.30

Robust z-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

(1) Standard errors are clustered at site level.

(2) Columns 2,4 (IV results) present results correcting for self-selection using being born in 2002 as an instrument.

(3) Lagged anthropometric indicators are instrumented throughout, including in columns marked OLS, using birth size and death of a household member during pregnancy as instruments.

(4) Base category: rural, female, Other Castes, Coastal AP, not drought-affected.

(5) Coefficients on male, caste, urban and region dummies, caregiver's education, wealth index and household size are not reported here.

Results from estimating equation (2) are presented in Table 3. Having suffered from drought between the two rounds has a significant negative impact on the weight-for-age and height-for-age z scores in all specifications; the negative effect of drought on height-for-age is greater than on weight-for-age, which is consistent with the former being a more long term measure of health and thus capturing the effect of past health deprivation. In all four regressions the MDMS dummy has a positive impact; this effect is significant at the 1% level in both the un-instrumented and instrumented results for weight-for-age and for the un-instrumented results

at home and without medical attention. It is important to note that our results do not depend on the inclusion or instrumenting of the lagged dependent variable. The patterns around the impact of the drought, and the cushioning effect of the Midday Meals, are similar in sign and statistical significance (although with even greater magnitudes) if we redefine our estimated equation using changes in z scores as the outcome variable and omit the lagged dependent variable from the regressors.

on height-for-age (but not when selection into treatment is accounted for). Standard errors in all regressions are clustered at the sub-district level. Thus, according to the preferred IV specifications, the Midday Meal Scheme increased weight-for-age by 0.60 s.d. and increased height-for-age by 0.27 s.d. (and not statistically significant), which is consistent with larger short-term impacts of changes in nutritional intake on weight.

These results, while documenting health deprivation as a result of the drought, do not provide any direct evidence of whether school meals are able to compensate for the effect of the past droughts in our sample. To answer this question, we next present the main estimation results based on estimation of equation (3), which allows for heterogeneous impacts of school meals on children whose households had suffered from drought between the two rounds. Table 4 presents the resulting estimates both with and without correcting for self-selection into the treatment.

Table 4: Heterogeneous impact of Midday meals on drought-affected children's nutrition

VARIABLES	(1)	(2)	(3)	(4)
	Weight-for-age in 2006/7 OLS	IV	Height-for-age in 2006/7 OLS	IV
Midday Meals	0.068* (1.67)	0.31 (1.31)	0.15** (1.98)	-0.17 (-0.52)
Drought	-0.23*** (-3.43)	-0.44*** (-4.00)	-0.33*** (-5.71)	-0.77*** (-3.86)
MDMS x Drought	0.21*** (2.81)	0.62*** (2.73)	0.19*** (3.21)	0.98** (2.39)
Age expressed in years	-0.091 (-1.12)	-0.32*** (-2.69)	0.38*** (2.71)	0.38* (1.73)
Weight-for-age in R1	0.61*** (7.62)	0.58*** (7.89)		
Height-for-age in R1			0.58*** (5.25)	0.58*** (5.70)
Constant	-0.49 (-1.33)	0.55 (1.03)	-2.96*** (-4.54)	-2.81*** (-2.82)
Observations	1,199	1,199	1,178	1,178
R-squared	0.380	0.339	0.219	0.180
Kleibergen-Paap F-statistic	31.7	8.00	11.4	8.15
Hansen J-statistic p value	0.71	0.27	0.36	0.17

Robust z-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

(1) Standard errors are clustered at site level.

(2) Columns 2,4 (IV results) present results correcting for self-selection using being born in 2002 as an instrument.

(3) Lagged anthropometric indicators are instrumented throughout, including in columns marked OLS, using birth size and death of a household member during pregnancy as instruments.

(4) Base category: rural, female, Other Castes, Coastal AP, not drought-affected.

(5) Coefficients on male, caste, urban and region dummies, caregiver's education, wealth index and household size are not reported here.

As can be seen, having suffered from drought in the past four years has a significant negative impact on both height-for-age and weight-for-age across all specifications. However, the negative impact of drought is compensated for by school-feeding in all specifications. The over-identification tests for the IV regressions fail to reject the null of all instruments being exogenous. Correcting for self-selection, the estimates of both the negative impact of the drought and the effect of school-feeding on drought-affected children rise substantially. The compensatory effect of the Midday Meals is statistically significant across all selection-corrected estimates at the 5% level of significance. Results in this table suggest that the positive effect of the scheme on average, as documented in Table 3, is concentrated

among drought-affected children in the sample. For weight-for-age, the impact of the program on children not experiencing drought is positive but not statistically significant; for height-for-age the effect is negative and statistically insignificant. The heterogeneity in the effect of the school meals, thus, appears to be central to understanding the health benefits of the Scheme.

The positive effect of the midday meals is larger for both health measures, across all specifications, than the negative impact of the drought, indicating that school meals more than compensate for the negative impact of the drought. However, the overcompensation effect is not statistically significant as F-tests investigating whether the sum of coefficients of Drought and its interaction with MDMS is different from zero are not able to reject the null in most specifications. This pattern is also true of other ways of measuring drought where also the null cannot be rejected.

One potential cause for concern in interpreting our estimates is that our drought measure is a self-reported binary variable which equals one if a household reports having suffered from drought in the past four years (i.e. between the two survey rounds) and zero otherwise. There could be systematic reporting bias in this variable that is correlated with time-varying unobservables which affect changes in nutrition. We do not think this likely to be a severe problem given that the mean incidence of drought does not differ significantly at the 5% level between our treatment and comparison groups. Nonetheless, as a robustness check we re-ran our estimation using village-level averages of reported drought instead of self-reported drought; results from this exercise are shown in Appendix 3

and display a very close similarity in the pattern of incidence of benefits from the Scheme.

To avoid self-reporting bias, we can also use reports of natural disasters from the community questionnaires collected at the same time as the household data. A further advantage of using data from the community questionnaire is that, unlike the household questionnaire, we have information on the timing of droughts that affected the village in the last four years. This is important in order to assess whether the effect of the Midday Meals in cushioning the impact of drought is mostly contemporaneous (i.e. compensating for recent droughts) or whether it is compensating for health deterioration in the past (i.e. leading to catch-up growth). Context instruments were administered in each of the communities (villages or urban wards) from which the data are collected; these collected information from local key informants on the natural disasters that affected the community between rounds, including how long ago the disaster had taken place. 50 out of 101 communities reported having been affected by drought in the past four years, of which 19 reported the drought had happened in the last 13 months; all other communities reported the drought as having occurred at least 18 months ago.²⁹

We used this information to re-run our analysis in the following way: first using just the community-level variable for whether a drought had happened in the last four years instead of the self-reported drought measure, then only using a dummy variable for a drought in the last 13 months and finally only using a dummy variable for a drought at least 18

²⁹Three communities reported drought twice in the intervening period. We used the more recent drought from that community in the estimation.

Table 5: Results using community-level information on timing and occurrence of drought

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Drought happened in the last 4 years				Drought happened 18+ months ago				Drought happened in last 13 months			
	Weight-for-age		Height-for-age		Weight-for-age		Height-for-age		Weight-for-age		Height-for-age	
	(2007)		(2007)		(2007)		(2007)		(2007)		(2007)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Midday Meals	0.064	0.41	0.083	-0.15	0.11**	0.45**	0.14	0.0035	0.14***	0.67***	0.22***	0.35
	(1.11)	(1.50)	(0.87)	(-0.45)	(2.46)	(2.25)	(1.59)	(0.013)	(2.74)	(2.86)	(2.79)	(0.86)
Drought	-0.046	-0.10	-0.30***	-0.55***	-0.066	-0.23*	-0.25**	-0.57**	0.011	0.15	0.030	0.11
	(-0.69)	(-0.71)	(-2.95)	(-2.74)	(-0.77)	(-1.65)	(-2.14)	(-2.40)	(0.094)	(0.98)	(0.23)	(0.55)
MDMS x Drought	0.14**	0.29	0.21**	0.64**	0.083	0.45*	0.19*	0.79**	0.073	-0.19	-0.0088	-0.14
	(2.15)	(1.33)	(2.14)	(2.12)	(1.19)	(1.86)	(1.65)	(2.04)	(0.85)	(-0.90)	(-0.074)	(-0.50)
Age expressed in years	-0.085	-0.33***	0.41***	0.37*	-0.079	-0.33***	0.40***	0.37	-0.090	-0.35***	0.39***	0.33
	(-1.02)	(-2.68)	(3.00)	(1.78)	(-0.93)	(-2.74)	(2.88)	(1.62)	(-1.09)	(-2.67)	(2.93)	(1.47)
Weight-for-age in R1	0.61***	0.57***			0.61***	0.58***			0.61***	0.57***		
	(7.52)	(7.21)			(7.81)	(7.87)			(7.78)	(7.50)		
Height-for-age in R1			0.58***	0.56***			0.57***	0.57***			0.58***	0.57***
			(5.44)	(5.67)			(5.45)	(5.57)			(5.76)	(6.27)
Constant	-0.54	0.56	-3.05***	-2.75***	-0.58	0.56	-3.08***	-2.80**	-0.55	0.50	-3.12***	-2.89***
	(-1.42)	(0.98)	(-4.74)	(-2.83)	(-1.52)	(1.00)	(-4.68)	(-2.52)	(-1.47)	(0.84)	(-4.59)	(-2.71)
Observations	1,199	1,199	1,178	1,178	1,199	1,199	1,178	1,178	1,199	1,199	1,178	1,178
R-squared	0.376	0.342	0.211	0.220	0.374	0.318	0.219	0.194	0.377	0.328	0.204	0.216
Kleibergen-Paap F-statistic	35.1	7.13	11.6	10.0	37.5	8.86	13.0	11.0	39.4	6.05	13.1	12.2
Hansen J-statistic p value	0.91	0.46	0.40	0.31	0.93	0.56	0.47	0.53	0.86	0.49	0.46	0.38

Robust z-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

(1) Standard errors are clustered at site level.

(2) Columns 2,4, 6, 8, 10, 12 (IV results) present results correcting for self-selection using being born in 2002 as an instrument.

(3) Lagged anthropometric indicators are instrumented throughout, including in columns marked OLS, using birth size and death of a household member during pregnancy as instruments.

(4) Base category: rural, female, Other Castes, Coastal AP, not drought-affected.

(5) Coefficients on male, caste, urban and region dummies, caregiver's education, wealth index and household size are not reported here.

months ago. Results from this analysis are presented in Table 5. As can be seen, the effect of drought is negative (although not statistically significant for weight-for-age) in the first set of results which use a dummy variable for whether a drought had happened in the last four years and there is a significant positive impact for the Midday Meals across both measures of nutrition; this pattern breaks down entirely in the case where drought occurred in the last 13 months and coefficients on neither drought nor its interaction with Midday Meals are significant; finally, in the case where the drought happened at least 18 months ago, both the impact of the drought and the safety-net impact of the Midday meals are strongly significant and close in magnitude to our results using self-reported drought. Thus, it appears that Midday Meals are compensating for the negative effects of the severe 2002-3 drought rather than the more recent, less severe drought. As noted previously, we believe this result is a reflection of the much greater severity of the 2002-3 drought than any other droughts in the study period, and should not be taken as evidence that nutritional interventions have a greater role to play in addressing the effects of droughts only after some time has elapsed.

We also ran the analysis on a series of sub-samples to better understand the pattern of nutritional benefits. In particular, we re-ran the analysis sequentially restricting the sample to rural children and by whether they were stunted/underweight in Round 1 of the survey or not. We found that the results are driven entirely by the rural sample, and that they are driven entirely by children who were not stunted/underweight in 2002.³⁰

³⁰Results are not presented here in the interests of space. It should be noted that as a result of severely restricting sample sizes, occasionally the statistical significance of the results declines to 10% level of significance and F-statistics also go down.

This latter result, that the effect seems to be most conspicuous in the sub-sample of children who were not stunted/underweight in 2002, also agrees with the nutritional literature that finds that children stunted in the first years of life have more difficulty in achieving catch-up growth (Adair 1999; Crookston et al. 2010).

As described earlier in the discussion on identification (Section 4), one might be concerned that the positive interaction effect of drought and school meals (enrollment) is due to a nonlinear effect of the age of first exposure to drought. However, when we estimate our base specifications on the sample of children in rural areas who have enrolled in or plan to enrol in private schools, we do not find any effects of being born after December 2001 on health outcomes of drought affected children. This is true whether we estimate in IV regression with Born2002 as an instrument for enrollment, or estimate the direct effects of Born2002 interacted with Drought on child health outcomes after controlling linearly for Age and Age interacted with Drought.³¹

While this provides suggestive evidence that our exclusion restriction is not violated by non-linear effects of the age of exposure to drought, one might be still be concerned that parents of children enrolled in private school are wealthier than parents of children enrolled in public schools (and are thus able to smooth shocks better) and the non-result merely reflects this pattern. However, while households in the private schools sub-sample are wealthier on average, there is considerable overlap in the

³¹Our instrument is very weak when applied only to the sample of children who are already in or will soon attend private schools; this is because children about to enrol in private schools typically would spend a longer period (about two years) in pre-primary Kindergarten classes than children enrolling in government schools who make the transition from public preschools (anganwadis) to primary school earlier.

wealth index between households in the treatment group and the private school sub-sample in rural areas. Our non-result holds even when we restrict the sample to rural households with wealth levels similar to our treatment group. We also note that within rural areas the baseline wealth index did not differ significantly between the drought-affected households and the non-affected households even at the 10% level of significance, which reflects perhaps the severity of the 2002-3 drought.

Finally, in the light of a large literature documenting the long-term impacts of environmental shocks in early childhood (e.g. Almond 2006; the literature surveyed in Almond and Currie 2011), it seems implausible to us that the nonlinear effects of age of exposure to the drought could be such that older children (who were only two months older than the control sample) experienced no negative effects from the 2002-3 drought while children in the control sample experienced large negative effects of drought on nutrition. The sample children were nearly all less than two years of age when the drought hit. This is, however, precisely the claim that would have to be put forth if we are to believe that the pattern of Midday Meals compensating for the negative impacts of drought is, in fact, an artefact of direct non-linear impacts of age on nutrition.

As a final robustness check, we report results using the full sample rather than just children attending or planning to attend public schools. Results are reported in Appendix 4. The results from the IV specifications are substantially similar when using the full sample.³²

³²Although results using the full sample exhibit the same patterns in the sign and magnitude of the coefficients, these are not always statistically significant. The insignificance of our results at times in the full sample is a product of the IV that we use. Norms around the age when children may be admitted to school are much more implemented in the public schooling sector than in private institutions which are more

6 Conclusion

Midday Meals, as Afridi (2010) rigorously documents, represent a substantial increase in nutritional intake for children. Given that children in her study (with a mean age of 8.5 years) were about 3 to 4 years older on average than our treatment group, that official guidelines on the minimum nutritional content of school meals are not sensitive to age of the child, and that these guidelines were revised substantially upwards from the 2006-7 school year (whereas Afridi's data is from 2004), it is plausible in our opinion that the nutritional increment from the school meals program could be substantially greater even in comparison to the large increments that are documented in her study. Combined with the vulnerability of the children due to a severe drought in early childhood, and an extensive literature documenting that these negative effects can be quantitatively very large, we view the large cushioning impacts of the Midday Meals as plausible.

These results also seem to make intuitive sense: children in drought-stricken areas see a decline in nutritional intake impacting their health negatively, but the Midday Meals Scheme in these situations acts as a safety net compensating for this previous health shock, at least for young children just entering school. In a context where preschool nutritional programmes, most notably the Integrated Child Development Scheme, face major weakness in delivery and have not been able to universalize access such a role is important.³³

amenable to admitting students at younger ages as well. Thus our IV is much less informative in the full sample than it is in the restricted sample of children who are in or will later join public schools. This is borne out by weaker first-stage results and much lower F-statistics when using the full sample instead of the restricted sample.

³³Note that our results do not imply that preschool feeding would not be as effective or perhaps even more effective since it targets children at younger ages. This is an

Our findings in this paper resonate with the opinion of Alderman and Bundy (2012) who conclude in a recent review article that it is quite likely that “Food For Education is a plausible candidate for social protection on par with Conditional Cash Transfer Programmes”. Taken in conjunction with findings from the literature reported in Alderman and Bundy (2012), our results could be taken to support a broader program of individually targeted food transfers to young children in areas suffering from various shocks.

Finally, in discussing the wider applicability of these results, it should be noted that Andhra Pradesh is one of the better performers among Indian states in service delivery generally, and in the Midday Meals Scheme in particular. The superior performance of the program in AP has been noted in both the academic literature (e.g. Drèze and Goyal 2003) and administrative reviews of the Scheme (Saxena 2003). The findings may not generalize to other states within India, especially to states such as Uttar Pradesh and Bihar noted as poor implementers of the Scheme, unless the delivery mechanisms and political/administrative will can also be raised to similar levels.

The effect of school meals as a safety net can be of much importance. Much of India’s population depends on agriculture for their livelihood; agricultural shocks, of which droughts are the most prominent example in many parts of India including Andhra Pradesh, lead to a decline in household food availability and a worsening of child nutrition and health.

The pernicious impact of this childhood nutritional deprivation on an

important point, stressed in relation to school feeding in India and elsewhere by both Haddad (2011) and Alderman and Bundy (2011), which needs to be considered explicitly in deriving policy implications from this study.

individual's health and nutritional status may persist into adulthood, and is likely to affect their ability to function fully in daily life. If school meals can cushion children from these shocks and reduce the variability in intra-seasonal food intake of children, it may be of great importance for their future biological development. This effect of school meals has not, to our knowledge, been studied or highlighted at all in the academic literature but may be worth evaluating separately in future studies.

This omission in the academic literature regarding the role of school feeding in social protection is especially surprising given that the same is not true of related administrative and policy documents. Our findings indicate that the role of the safety net, at least for younger children, is very significant.

We believe that these results, combined with other evidence on the positive impact of school meals on school participation and daily nutrient intake, provide empirical support for the benefits of the program in India. With regard to the Indian context, this is one of the few attempts at a rigorous evaluation of a scheme that covers more than 120 million children nationally and as such its findings should be of obvious interest to administrators and policy makers working on health and education.

References

- ADAIR, L. S. (1999): "Filipino children exhibit catch-up growth from age 2 to 12 years," *The Journal of Nutrition*, 129(6), 1140–1148.
- AFRIDI, F. (2010): "Child welfare programs and child nutrition: Evidence from a mandated school meal program in India," *Journal of Development*

- Economics*, 92(2), 152–165.
- (2011): “The impact of school meals on school participation: Evidence from rural India,” *Journal of Development Studies*, 47(11), 1636–1656.
- AKRESH, R., P. VERWIMP, AND T. BUNDERVOET (2011): “Civil war, crop failure, and child stunting in Rwanda,” *Economic Development and Cultural Change*, 59(4), 777–810.
- ALDERMAN, H., AND D. BUNDY (2012): “School Feeding Programs and Development: Are We Framing the Question Correctly?,” *The World Bank Research Observer*, 27(2), 204–221.
- ALMOND, D. (2006): “Is the 1918 Influenza pandemic over? Long-term effects of in utero Influenza exposure in the post-1940 US population,” *Journal of Political Economy*, 114(4), 672–712.
- ALMOND, D., AND J. CURRIE (2011): “Killing me softly: The fetal origins hypothesis,” *The Journal of Economic Perspectives*, 25(3), 153–172.
- BEHRMAN, J. R., AND A. B. DEOLALIKAR (1988): “Health and nutrition,” *Handbook of Development Economics*, 1, 631–711.
- BEHRMAN, J. R., AND J. HODDINOTT (2005): “Programme evaluation with unobserved heterogeneity and selective implementation: The Mexican PROGRESA impact on child nutrition,” *Oxford Bulletin of Economics and Statistics*, 67(4), 547–569.
- BOERSMA, B., AND J. M. WIT (1997): “Catch-up growth,” *Endocrine Reviews*, 18(5), 646–661.

- COLY, A. N., J. MILET, A. DIALLO, T. NDIAYE, E. BÉNÉFICE, F. SIMONDON, S. WADE, AND K. B. SIMONDON (2006): “Preschool stunting, adolescent migration, catch-up growth, and adult height in young Senegalese men and women of rural origin,” *The Journal of Nutrition*, 136(9), 2412–2420.
- CROOKSTON, B. T., M. E. PENNY, S. C. ALDER, T. T. DICKERSON, R. M. MERRILL, J. B. STANFORD, C. A. PORUCZNIK, AND K. A. DEARDEN (2010): “Children who recover from early stunting and children who are not stunted demonstrate similar levels of cognition,” *The Journal of Nutrition*, 140(11), 1996–2001.
- DRÈZE, J., AND A. GOYAL (2003): “Future of Mid-Day Meals,” *Economic and Political Weekly*, 38(44), 4673–4683.
- FERREIRA, F. H., AND N. SCHADY (2009): “Aggregate economic shocks, child schooling, and child health,” *The World Bank Research Observer*, 24(2), 147–181.
- FINANCIAL EXPRESS (2003): “Cheer up, Drought may be over soon,” *Financial Express*, July 14.
- (2004): “2002 Revisited as Dry Spell continues,” *Financial Express*, July 22.
- FRONTLINE (2002): “The Ground Realities,” *Frontline Magazine*, Aug 17–30.
- GOLDEN, M. (1994): “Is complete catch-up possible for stunted malnourished children?,” *European Journal of Clinical Nutrition*, 48, S58–70.

- HADDAD, L. (2013): “Why India needs a national nutrition strategy,” *British Medical Journal*, 343, d6687.
- JACOBY, H. G. (2002): “Is there an intrahousehold flypaper effect? Evidence from a school feeding programme,” *The Economic Journal*, 112(476), 196–221.
- JAYARAMAN, R., AND D. SIMROTH (2011): “The impact of school lunches on primary school enrollment: Evidence from India’s midday meal scheme,” ESMT Working Paper 11-11, European School of Management and Technology, Berlin.
- KHERA, R. (2006): “Mid-day meals in primary schools: Achievements and challenges,” *Economic and Political Weekly*, pp. 4742–4750.
- KUMAR, D. (2002): “No Silver Lining,” *Outlook Magazine*, August 12.
- MACCINI, S., AND D. YANG (2009): “Under the weather: Health, schooling, and economic consequences of early-life rainfall,” *The American Economic Review*, 99(3), 1006–1026.
- MUKHERJEE, A., AND S. CHAKRABORTY (2004): “Drought, Ahoy?,” *Outlook Magazine*.
- SENAUER, B., AND M. GARCIA (1991): “Determinants of the nutrition and health status of preschool children: an analysis with longitudinal data,” *Economic Development and Cultural Change*, pp. 371–389.
- STRAUSS, J., AND D. THOMAS (2007): “Health over the life course,” *Handbook of Development Economics*, 4, 3375–3474.

- TANNER, J. M. (1981): “Catch-up growth in man,” *British Medical Bulletin*, 37(3), 233–238.
- TATA, P. (2002): “Drought hits India,” *New Scientist*, August 03.
- THE HINDU (2003): “Not a Spectacular Recovery,” *The Hindu* October 3.
- THORAT, S., AND J. LEE (2005): “Caste discrimination and food security programmes,” *Economic and Political Weekly*, 40(39), 4198–4201.
- TIMES OF INDIA (2003): “Poor Rabi Harvest Likely after disappointing Kharif,” New Delhi April 2.
- VERMEERSCH, C., AND M. KREMER (2004): “School meals, educational Attainment, and school competition: evidence from a randomized evaluation,” World Bank Policy Research Working Paper 3523, World Bank, Washington, D.C.

Appendices for Chapter 3

Table A.1: Descriptive statistics for treatment group, comparison group and all non-beneficiaries

VARIABLES	Treatment Group	Restricted Comparison Group	All non-beneficiaries	Total
Male	0.511 (0.5)	0.506 (0.5)	0.551 (0.498)	0.537 (0.499)
Urban	0.058 (0.233)	0.103*** (0.304)	0.347*** (0.476)	0.244 (0.43)
Drought	0.34 (0.474)	0.386* (0.487)	0.244*** (0.43)	0.278 (0.448)
Wealth index (2002)	0.327 (0.161)	0.312 (0.16)	0.45*** (0.21)	0.406 (0.203)
Scheduled Castes	0.23 (0.421)	0.211 (0.408)	0.158** (0.365)	0.184 (0.387)
Scheduled Tribes	0.212 (0.409)	0.164** (0.371)	0.112*** (0.315)	0.147 (0.354)
Backward Classes	0.439 (0.497)	0.515** (0.5)	0.479 (0.5)	0.465 (0.499)
Other Castes	0.119 (0.325)	0.11 (0.313)	0.252*** (0.434)	0.205 (0.404)
Telangana region	0.279 (0.449)	0.375*** (0.485)	0.389*** (0.488)	0.35 (0.477)
Rayalaseema region	0.344 (0.475)	0.295* (0.456)	0.277*** (0.448)	0.301 (0.459)
Coastal Andhra Pradesh	0.377 (0.485)	0.33* (0.471)	0.334* (0.472)	0.349 (0.477)
Height-for-age z score (2002)	-1.351 (1.461)	-1.597*** (1.53)	-1.27 (1.487)	-1.298 (1.478)
Weight-for-age z-score (2002)	-1.621 (1.064)	-1.843*** (1.167)	-1.504** (1.158)	-1.546 (1.127)
Height-for-age z-score(2007)	-1.645 (0.83)	-2.1*** (0.96)	-1.66 (1.068)	-1.655 (0.989)
Weight-for-age z-score (2007)	-1.879 (0.854)	-2.181*** (0.865)	-1.859 (0.977)	-1.866 (0.935)
Age (in years)	5.5 (0.276)	5.29*** (0.324)	5.343*** (0.334)	5.399 (0.323)
N	695	536	1255	1950

Table A.2: First stage results for endogenous variables

VARIABLES	MDMS	MDMS x Drought	Weight-for-age z score (R1)	Height-for-age z score (R1)
Born in 2002	-0.25*** (-4.67)	0.044 (1.68)	-0.042 (-0.26)	-0.13 (-0.65)
Born2002 x Drought	-0.047 (-0.76)	-0.42*** (-9.02)	0.088 (0.74)	0.48* (1.96)
perception of child's size at birth	-0.018* (-1.76)	-0.0049 (-0.55)	-0.29*** (-7.70)	-0.21*** (-3.31)
death of household members	-0.19* (-2.05)	-0.16* (-1.82)	-0.39 (-1.65)	-0.35 (-1.63)
in last 4yrs has hh suffered drought	-0.017 (-0.50)	0.68*** (18.0)	0.052 (0.79)	0.091 (0.58)
Age expressed in years	0.23** (2.77)	0.080 (1.56)	-0.56** (-2.54)	-0.72** (-2.53)
Constant	-0.53 (-1.13)	-0.40 (-1.44)	2.75** (2.20)	3.29** (2.16)
Observations	1,215	1,215	1,200	1,184
R-squared	0.174	0.521	0.134	0.167

Robust z-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

(1) Standard errors are clustered at site level.

(2) Columns 2,4 (IV results) present results correcting for self-selection using being born in 2002 as an instrument.

(3) Lagged anthropometric indicators are instrumented throughout, including in columns marked OLS, using birth size and death of a household member during pregnancy as instruments.

(4) Base category: rural, female, Other Castes, Coastal AP, not drought-affected.

(5) Coefficients on male, caste, urban and region dummies, caregiver's education, wealth index and household

Table A.3: Results using site-averaged drought measure

VARIABLES	(1) Weight-for-age	(2) Weight-for-age (2007)	(3)	(4) Height-for-age (2007)
Midday Meals	0.057 (1.29)	0.39 (1.61)	0.090 (1.01)	-0.18 (-0.52)
MDMS x Drought (village-average)	0.24**	0.52	0.30**	1.00
Drought (village-average)	(2.00) -0.33*** (-2.74)	(1.59) -0.40* (-1.94)	(2.01) -0.90*** (-6.84)	(1.57) -1.26*** (-4.05)
Age expressed in years	-0.082 (-1.03)	-0.32*** (-2.62)	0.39*** (2.93)	0.40* (1.84)
Weight-for-age in R1	0.61*** (7.43)	0.57*** (7.22)		
Height-for-age in R1			0.57*** (5.22)	0.56*** (5.52)
Constant	-0.53 (-1.48)	0.55 (0.99)	-2.98*** (-5.09)	-2.85*** (-3.00)
Observations	1,199	1,199	1,178	1,178
R-squared	0.379	0.345	0.249	0.248
Kleibergen-Paap F-statistic	33.5	8.05	11.1	8.47
Hansen J-statistic p value	0.72	0.33	0.21	0.12

Robust z-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

(1) Standard errors are clustered at site level.

(2) Columns 2,4 (IV results) present results correcting for self-selection using being born in 2002 as an instrument.

(3) Lagged anthropometric indicators are instrumented throughout, including in columns marked OLS, using birth size and death of a household member during pregnancy as instruments.

(4) Base category: rural, female, Other Castes, Coastal AP, not drought-affected.

(5) Coefficients on male, caste, urban and region dummies, caregiver's education, wealth index and household

Table A.4: Estimates using the whole sample

VARIABLES	(1) Weight-for-age OLS	(2) (2007) IV	(3) OLS	(4) Height-for-age (2007) IV
Midday Meals	0.063* (1.78)	0.70 (1.51)	0.070 (1.39)	0.10 (0.23)
Drought	-0.16*** (-4.38)	-0.24 (-1.51)	-0.28*** (-4.15)	-0.57*** (-2.99)
MDMS x Drought	0.14*** (2.88)	0.30 (0.80)	0.14** (2.32)	0.80* (1.70)
Age expressed in years	-0.056 (-1.15)	-0.30** (-1.97)	0.40*** (4.79)	0.30* (1.81)
wealth index	0.34*** (2.62)	0.53*** (3.23)	0.39* (1.85)	0.43** (2.36)
Weight-for-age in R1	0.65*** (11.0)	0.62*** (10.1)		
Height-for-age in R1			0.57*** (6.96)	0.55*** (7.25)
Constant	-0.68** (-2.24)	0.27 (0.40)	-3.09*** (-6.87)	-2.57*** (-3.57)
Observations	1,900	1,900	1,874	1,874
R-squared	0.402	0.309	0.266	0.248
Kleibergen-Paap F-statistic	50.7	5.11	24.3	4.36
Hansen J-statistic p value	0.86	0.67	0.25	0.13

Robust z-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

(1) Standard errors are clustered at site level.

(2) Columns 2,4 (IV results) present results correcting for self-selection using being born in 2002 as an instrument.

(3) Lagged anthropometric indicators are instrumented throughout, including in columns marked OLS, using birth size and death of a household member during pregnancy as instruments.

(4) Base category: rural, female, Other Castes, Coastal AP, not drought-affected.

(5) Coefficients on male, caste, urban and region dummies, caregiver's education, wealth index and household