The impact of child work on cognitive development: results from four Low to Middle Income countries

IFS Working Paper W18/29

Michael Keane

Sonya Krutikova

Timothy Neal

The Impact of Child Work on Cognitive Development: Results from Four Low to Middle Income Countries

by

Michael Keane[1]

Sonya Krutikova[2]

Timothy Neal[3]

August 17, 2018
Revised November 1, 2018

Abstract: We study the relationship between child work and cognitive development in four Low and Middle Income Countries. We address a key weakness in the literature by including children's full time-use vector in the analysis, which leads to different findings from previous studies which do not distinguish between alternative counter-factual activities. We find child work is only detrimental if it crowds out school/study time rather than leisure. Furthermore, the marginal effects of substituting domestic chores or economic activities for school/study time are similar. Thus, policies to enhance child development should target a shift from all forms of work toward educational activities.

JEL Codes: I25, J13, J24, O15

Keywords: Child labour; Child development; Education; Time use; Item response theory; Value added models

Acknowledgements: We thank seminar participants at Essex, Oxford and the IFS. Keane and Neal's work on this project was supported by Australian Research Council (ARC) grants FL110100247, DP1094719 and CE110001029. Krutikova gratefully acknowledges the support of the Economic and Social Research Council's Centre for the Microeconomic Analysis of Public Policy at the IFS (grant reference ES/M010147/1).

[1] University of New South Wales, Australia: m.keane@unsw.edu.au

[2] Institute for Fiscal Studies, UK: sonya_k@ifs.org.uk

[3] University of New South Wales. Australia: timothy.neal@unsw.edu.au

I. Introduction

Is child labour harmful for child development and human capital accumulation? The evidence to date suggests the answer is "probably yes." But the evidence is still inconclusive and inadequate for directing policy efforts. A critical gap which has received little attention is an understanding of the trade-offs that children face between working and other ways of spending their time (e.g., school, leisure, chores). Existing studies typically estimate models which only include indicators of child work, and do not control for how children spend their non-work time. Even if consistent, such estimates of the trade-off between working and a counterfactual *bundle* of activities is of limited use for directing policy for two reasons. First, it only informs us about what would happen if the time freed up through working less (or not at all) was reallocated to the other activities in the same proportions as in the current bundle. Second, it is not informative about the effects of alternative policies that aim to shift work time to a *specific* time use alternative, such as enhanced school time or enhanced leisure time.

We address this gap in the literature by estimating child cognitive ability production functions that account for the complete time budget of children – including time spent in paid and unpaid work, doing chores and attending school. This allows us to generate important new evidence on the impact of child work[4] on human capital accumulation. In particular, by studying trade-offs between working and each of the other ways in which children spend their time, we go beyond aggregated estimates of the impact of child work, to identify the more and less productive counterfactual time-use activities. We show clearly that, irrespective of context and child age, the answer to the question of whether child work is harmful depends crucially on what the child would do instead of working. Policies to reduce child work will only lead to gains in human capital if they incentivize families to reallocate the freed up time to the subset of possible alternative activities that are more productive than working.

Our analysis is made possible by the Young Lives data collected since 2002. Young Lives is a unique multi-country data-set that contains unusually rich panel data on (close to) nationally representative samples of children followed from birth until age 22 in four low and middle income countries – Ethiopia, India, Vietnam and Peru. These data record information on a complete vector of activities children engage in during a "typical" 24 hour period, including time spent going to school, studying at home, doing paid work, chores and farm

[4] We refer to "child work" rather than "child labour" as child labour tends to be defined as a narrower set of activities which excludes domestic chores, for example (e.g. International Labour Organization 2017). Since the evidence and theoretical justification for focusing on a subset of work activities (e.g. market work only) is inconclusive, we consider all work-related activities undertaken by children including domestic chores and care responsibilities in the household, farm tasks and work on family business, as well as paid work outside the household. Hence we adopt the broader term "child work" to describe all these activities.

work, caring for others at home, sleeping and playing. This allows us to study trade-offs among time spent on all these activities for accumulation of human capital. Furthermore, the data uses state-of-the-art methods to measure children's skills using scores on mathematics and verbal tests administered in four out of five survey rounds to the full sample of children.

We utilize the Young Lives data to estimate child cognitive ability production functions that control for a complete vector of time inputs as well as a rich set of controls for other inputs (i.e., parent education, wealth, household structure, demographics). Of course, it is not feasible to measure all possible inputs to child development, or to control perfectly for unobserved initial ability. Thus, we utilise the panel dimension of the data to estimate value added (VA) production functions in which omitted inputs and unobserved ability are assumed to be captured by a lagged achievement measure (Todd and Wolpin 2007, Fiorini and Keane, 2014). As with all econometric methods, the validity of our approach hinges on assumptions which cannot be directly tested. However, a group of studies in the education literature offer encouraging evidence on the effectiveness of the lagged test score as a control for unobserved heterogeneity, both through simulation (Guarino, Reckase, and Wooldridge 2014) and through comparisons of value added and experimental estimates (Angrist, Pathak, and Walters 2013; Deming et al. 2014; Muralidharan and Venkatesh 2013).

A striking finding is that leisure time is no more or less productive for child cognitive development than child work (including agricultural and paid work, as well as chores in the household). The implication is that policies that merely shift work time to leisure time will not enhance child cognitive development. In contrast, we find that school and study time are far more productive than child work. Thus, if one's goal is to enhance child development, it is important that policies aimed at reducing child work also target a shift in the time allocation toward educational activities.

We also show how the typical or "status quo" approach in the literature of estimating effects of child labour relative to a bundled counterfactual set of activities can generate misleading conclusions, particularly when different modes of reducing child labour result in a reallocation of time toward different bundles of non-work activities. As an important example of this point, the prior literature typically finds that work outside the household is much more harmful for child development than household work and chores. We find that this is largely due to the fact that work outside the household tends to crowd out school time, while household work tends to crowd out leisure time. In fact, we find that household work is almost as detrimental relative to school time as work outside the household – a result that has been masked in prior analysis.

The outline of the paper is as follows: Section II provides further background and review of the extensive literature on child labour. Section III describes the Young Lives data and Section IV describes our econometric methods. Section V presents our results and Section VI concludes.

II. Background and Literature Review

A large proportion of the world's children are engaged in some form of work. According to the International Labour Organisation (ILO) one in ten of all children in the world today (or 152 million) participate in what they define as "child labour."[5] Prevalence is highest in Africa where 20% of children are engaged in child labour, most of which takes place in agriculture (with 17% in services and 12% in the industrial sector). The numbers are higher still if participation in household chores is included. For instance, 54 million 5-14 year olds spend at least 21 hours a week on household chores – a threshold beyond which, descriptive evidence suggests, children struggle to combine work and school (ILO 2017).

The prevalence of child labour is among the most high profile policy issues facing Lower and Middle Income Countries (LMIC). Within policy circles there is a very broad consensus that child labour is detrimental to child development. Thus, the UN Sustainable Development Goals call for the eradication of all forms of child labour by 2025 (Target 8.7). According to the ILO, in the period between 2004 and 2014, 57 LMICs implemented a total of 279 specific policies, plans and programmes aimed at reducing the prevalence of child labour. The great majority of the world's children live in countries that have ratified the ILO's two main child labour conventions (ILO 2017).[6] In addition, there has been growing international pressure on companies to ensure child labour is not used at any stage of their production process; an example is the debate in the US over the Child Labor Deterrence Act which proposes to disallow import of goods produced using hazardous forms of child labour.[7]

Underlying the broad policy consensus for the eradication of child labour is the presumption that working is harmful for children. Given the widespread acceptance of this

[5] The ILO defines child labour as "work that is harmful to children's physical and mental development which includes work that is: mentally, physically, socially or morally dangerous and harmful to children; and which interferes with their schooling by: depriving them of the opportunity to attend school; obliging them to leave school prematurely; or requiring them to attempt to combine school attendance with excessively long and heavy work." (http://www.ilo.org/ipec/facts/lang--en/index.htm).

[6] Minimum Age Convention of 1973 (No. 138) and Worst Forms of Child Labour Convention, 1999 (No.182)

[7] Although the bill spurred intensive debate about the costs and benefits of banning child labour, and caused mass layoffs of children working in the garment industry in Bangladesh (Rahman, Khanam, and Absar 1999), it was never passed. It was not until 2015 that import of goods made by forced labour including forced or indentured child labour was officially banned in the Trade Facilitation and Trade Enforcement Act of 2015.

view, it is perhaps surprising that the academic literature on the effect of child labour on child development has not reached definitive conclusions. There are several sources of ambiguity:

Much of the existing evidence on human capital effects of child labour relies on school participation and progression measures as the outcome (Assad, Levison, and Dang 2010; Buonomo Zabaleta 2011; Patrinos and Psacharopoulos 1997; Ray and Lancaster 2003; Sedlacek et al. 2009). The findings in these studies are somewhat mixed. While the majority report at least moderate negative impacts of child labour on schooling, there are some studies suggesting that child labour and schooling could be complementary activities (Patrinos and Psacharopoulos 1997; Ravallion and Wodon 2000), contributing positively to schooling and study time below a certain threshold of intensity (Ray and Lancaster 2003).

The weakness of school participation and progression as measures of human capital has long been recognised in the literature. They are weaker predictors of adult income than direct measures of skills and proficiency (Glewwe 2002; Hanushek and Kimko 2000; Hanushek and Zhang 2009), and within the production function framework schooling is more suitably interpreted as an input (rather than an output).

There is some, though still limited, evidence on the impact of child labour on direct measures of human capital, such as test scores. Earlier studies found negative correlations between working and test scores across several contexts (Akabayashi and Psacharopoulos 1999; Heady 2003). More recent studies attempt to establish this relationship causally. Most find a negative relationship between work and attainment (Gunnarsson, Orazem, and Sánchez 2006; Bezerra, Kassouf, and Kuenning 2009; Emerson, Ponczek, and Souza 2017). Emerson et al. (2017) provides perhaps the most convincing evidence available to date of a negative relationship, using an individual fixed effects estimation strategy to show that working children have lower test scores by an amount equivalent to knowledge gained over roughly one quarter to three fifths of a year. The evidence is not conclusive, however. For example, (Dumas 2012)) finds that, for children in Senegal, participation in economic activities is not associated with lower cognitive achievement up to a threshold of 17 hours per week.

Further, there are some important limitations to existing studies based on test scores. First, these studies typically only have test score data for the subsample of children who are actually at school. This is clearly a selected sample, likely to leave out children most affected by child labour. Second, applying panel methods to study evolution of test scores over time (as in Emerson et al. 2017) requires tests that are comparable over time (Das and Zajonc 2010). But as far as we are aware, no existing studies claim to be using tests that are comparable over time.

Two key advantages of the Young Lives data that we draw on are that (i) it gathers test scores for all children irrespective of whether they are at school, and (ii) it provides comparable tests designed to be linkable to a common metric over time and across contexts.

As we stressed in the introduction, another limitation of existing child labour studies is that they fail to account for the complete time budget of children. In particular, the literature has a strong focus on the trade-off between work and schooling, but on the whole has little to say about trade-offs with other types of time-use, such as leisure/play or time spent studying at home. A typical underlying framework considers time-constraints given by time spent at school and working (e.g. Orazem and Gunnarsson 2004), while leisure and other activities are ignored. While there are more nuanced models which consider a more complete time-allocation vector (Cigno and Rosati 2005; Edmonds 2007), time inputs other than schooling and work have largely been absent from the empirical child labour literature.[8]

This approach is inconsistent with various pieces of evidence which suggest that the trade-off between child work and schooling is not one-to-one. For instance, Eric Edmonds uses the multi-country data from UNICEF's MICS project to show that across many contexts, up to a certain threshold hours of work increase without any or only slight change in school attendance (Edmonds 2007). In addition, there is evidence that observed decreases in child labour in response to conditional cash transfer programmes which reduce the relative price of schooling are much smaller than the increases in schooling suggesting that substitution away from leisure is also at play (Attanasio et al. 2010; Ravallion and Wodon 2000).

Clearly then schooling is unlikely to be the only activity that is crowded out by child labour (and vice-versa), and opportunity costs of work in terms of activities other than school attendance is a key missing piece in the study of the impact of child labour on human capital.[9] This is a gap that we are able to address using our unusually rich time-allocation data.

The Young Lives data provide rich measures of actual hours of child work, which are broken down into time spent on several activities including market work, farm work, work in

[8] There are a few exceptions - for instance Akabayashi and Psacharopoulos 1999.

[9] Fiorini and Keane (2014) emphasized the importance of accounting for the complete time use budget when estimating the effect of child care on child development in a developed country setting. We illustrate their point in the child labour context with a simple example. Assume a child's total time (T) is allocated between working (T_{wi}), leisure (T_{li}) and school (T_{si}), and that time not working is split equally between leisure and school, so $T_{li} = T_{si} = (T-T_{wi})/2$. Let the test score production function be given by $Y_i=\beta_0 + \beta_w T_{wi} + \beta_l T_{li} + \beta_s T_{si} + \varepsilon_i$, where $\beta_s = 2$ while $\beta_w = \beta_l = 0$. Thus, school has a positive effect on scores, while leisure and work have no effects. If we regress the test score Y_i on work time alone (without controlling for leisure and school time) we will obtain an estimated effect of (1/2)(-2) = -1. Thus, we find a negative effect of child work, despite the fact that a negative effect only exists if work crowds out school, not if it crowds out leisure. Further, our estimate of -1 *underestimates* the potential to improve scores by shifting time from child work to school, which actually has a positive effect of 2 per extra unit of school time.

the family business, and chores at home. Many studies focus on binary participation measures rather than the intensity of work (e.g., Psacharopoulos 1997; Ravallion and Wodon 2000; Emerson, Ponczek, and Souza 2017), but we are able to study child work effects on the intensive margin, accounting for time spent on all work-related activities. Also, many studies focus only on market work (Beegle, Dehejia, and Gatti 2009; Gunnarsson et al. 2006). However, (Edmonds 2007)) descriptive analysis of MICS data clearly suggests that what is likely to matter most for human capital accumulation is how long a child works, and that there is no empirical (or theoretical) justification for a sole focus on market activities.[10] In addition to including both types of activities in our analysis, we can test whether different types of work have different effects on child development.

Another key challenge in studying the effects of child labour is the endogeneity of children's time use. How much children work is one of a bundle of key household decisions. Thus, associations between child work and human capital may be confounded by omitted variable bias (e.g., omitted ability), as well as feedback from child outcomes to time inputs, and measurement error in inputs. It is difficult to imagine a feasible experimental design for testing the impact of child labour (as has been possible with some other material and time inputs into the human capital production function).[11] Studies that have attempted to move beyond correlational analysis have mostly relied on a range of instrumenting strategies, including instruments based on variation in child labour and school laws (Gunnarsson et al. 2006), gradual roll-out of conditional cash transfer programmes with school participation conditionality (e.g., Ravallion and Wodon 2000) and adverse weather, agricultural production and income shocks (Beegle et al. 2009; Jacoby and Skoufias 1997).[12]

Over a decade ago, in their overview of the child labour literature, Orazem and Gunnarson suggested that what would be needed to improve on the evidence are panel data with better repeated measures of human capital, which would allow one to apply the "value added" approach (Orazem and Gunnarsson, 2004). To date the literature has made limited progress in this direction. A handful of papers have utilised panel/recall data to estimate models which control for unobserved ability through lagged measures of human capital (Buonomo Zabaleta 2011; Dumas 2012) or individual fixed effects (as in the Emerson et al.

[10] There is little to suggest that market work is substantially more detrimental than domestic work (Edmonds 2007). Indeed, children engaged in paid work may be more likely to benefit from increased bargaining power within the household compared to those doing unpaid work or chores.
[11] For example interventions to improve key inputs such as nutrition and quality of parenting in early childhood (e.g. Rao et al. 2014) or quality of schooling in later childhood (Glewwe and Kremer 2006).
[12] Given how inextricably linked the child labour decision is to other key child and adult time-allocation decisions within the household, it is difficult to think of convincing instruments.

2017 paper we dicussed earlier). We build on this strand of work by exploiting the panel aspect of the Young Lives data to estimate value added models. Furthermore, we implement an “extended” value added model (suggest by Todd and Wolpin 2007) that includes not only lagged test scores but also lagged inputs. This relaxes some of the restrictive assumptions about the underlying technology of human capital production implied by the simple value added model (see Section IV for details).

An additional issue in prior literature is limited external validity, due to the ad hoc selection of ages and contexts for which evidence is available. Child work means very different things in different contexts and growing evidence on child development suggests that the same input can have different effects at different ages (Attanasio et al. 2017; Del Boca, Monfardini, and Nicoletti 2016). Our data allow us to look at effects across different contexts and ages but using comparable measures.

Finally, even if child labour is detrimental for child development *ceteris paribus*, it may still be an important source of household income, creating ambiguity about its overall effect. Child labour bans may still be harmful for children if they reduce household resources available to children (Basu 1999; Jacoby and Skoufias 1997). Thus, a complete accounting of the effect of child labour on child developments requires not only controlling for the complete time budget, but also for family income and measures of goods inputs into child quality production. To our knowledge this has not been done in prior work. The Young Lives data is not ideal in this dimension either, but at least it contains measures of household wealth and parent education that can be used to proxy for family income and goods inputs.

III. The Young Lives Data

The Young Lives data-set contains information on two cohorts of children from four LMIC countries: Ethiopia, India (Andhra Pradesh and Telengana), Peru and Vietnam. One cohort was born in 1994/95 while the other was born in 2001/02. Henceforth, we refer to these as the “Older Cohort” (OC) and “Younger Cohort” (YC), respectively. Both cohorts were surveyed in 2002, 2006, 2009 and 2013.[13] Thus, the YC is covered from birth roughly through age 12, while the OC is covered from rough ages 8/9 through ages 19/20.

Across the four countries the sample consists of 12,000 children. In each country, the 2,000 YC and 1,000 OC children were randomly sampled by selecting 100 YC and 50 OC children from each of 20 sites.[14] The data are statistically representative of the site level

[13] Another round of data was collected in 2016, but at the time of writing it was not yet publicly available.
[14] In Peru only 714 OC children were enrolled due to capacity constraints during the first round.

population and the 20 sites in each country were purposively selected to represent each country's socio-economic and geographic diversity, with a pro-poor bias.[15]

In this study we utilise data from the second, third and fourth data-collection rounds 2006, 2009 and 2013 for both the YC and the OC. Table 1 shows the average age at interview of each cohort in each of the four countries in the data collection rounds relevant for this paper. The average age at which children from the two cohorts were interviewed at each round is very similar. YC children are on average age 5, 8 and 12 in rounds 2, 3 and 4, respectively; the OC children are 12, 15 and 19. In each round, an extensive effort was made to find and interview children who had moved from their location in the previous survey round. As a result attrition between 2002 and 2013 across the four study countries is very low, ranging from 3.6% for the YC in Vietnam to 11% for the OC in Peru and Vietnam.[16]

Table 1: Average age at interview of Young Lives children

	Ethiopia		India		Peru		Vietnam	
	mean	sd	mean	sd	mean	sd	mean	sd
R2, 2006 (YC, 5)	5.198	0.316	5.395	0.309	5.332	0.390	5.306	0.302
R3, 2009 (YC, 8)	8.275	0.222	8.039	0.292	7.953	0.300	8.126	0.293
R4, 2013 (YC, 12)	12.122	0.324	11.986	0.316	11.912	0.305	12.195	0.308
R2, 2006 (OC, 12)	12.101	0.310	12.375	0.340	12.349	0.436	12.303	0.315
R3, 2009 (OC, 15)	15.028	0.295	14.980	0.339	14.934	0.340	15.091	0.320
R4, 2013 (OC, 19)	19.056	0.327	18.993	0.344	18.920	0.371	19.264	0.350

All data were collected through home visits by interviewers; these involve interviews with the person who knows most about the socio-economic circumstances of the household, the primary caregiver of the target child, as well as direct assessment of the target child. At older ages the target child was also interviewed. Young Lives was designed as a multi-purpose study of child poverty which is reflected in the breadth of the information collected during the interviews, including detailed data on household economic circumstances and demographic characteristics, primary caregiver background, role in the household, own socio-emotional well-being and aspirations for the target child, as well as detailed information about the education, health and development of the target child.

Several key variables for our analysis – child time-use and verbal and mathematical test scores – were not collected in the first round (2002) so we restrict our analysis to data

[15] With the exception of Peru, samples were not selected to be statistically representative. However, subsequent comparisons to nationally representative surveys in each country have shown that the data reflect the diversity of children in the two cohorts across a wide number of variables. See technical notes at www.younglives.org.uk.
[16] 6.3% YC Ethiopia; 9.2% OC Ethiopia; 4.8% YC India; 5.6% OC India; 7.3% YC Peru, 11.1% OC Peru; 3.6% in Vietnam, 11.3% OC in Vietnam.

from rounds 2, 3 and 4. Combining the two cohorts, we study outcomes at ages 8, 12, 15 and 19. Table 2 shows the total (N) and analysis sample sizes for each country at each of these ages. In most cases the analysis sample is slightly smaller than N due to missing values for the math or verbal assessments and/or core controls. However, there are instances where the analysis sample is much smaller than the full sample, for the following two reasons:

First, the verbal skills measure is language specific. Thus, in each country, we focus on the largest single language group. Ethiopia is the most multi-lingual country, and we focus on the children speaking Amharic, which is only about a quarter of the full sample. We also lose about 20% of the India sample for this reason, but Peru and Vietnam are little affected.

Second, we control for lagged time allocation variables in all specifications. But the time data were only collected for children age 5 and older, and not all of the children in the sample had turned 5 at the time of the Round 2 survey. This means we lose some of the YC age 8 observations. We lose about 10% for India, Peru and Vietnam, but we lose about 20% in Ethiopia (where the YC children were on average slightly younger than in the other three countries at the time of the round 2 survey – see Table 1).

Table 2: Analysis sample sizes by country, round and outcome variable

		Ethiopia			India			Peru			Vietnam	
	Math	Verbal	N	Math	Verbal	N	Math	Verbal	N	Math	Verbal	N
R3, 2009 (YC, 8)	1240	488	1885	1673	1383	1931	1846	1738	1942	1701	1640	1961
R4, 2013 (YC, 12)	1799	714	1867	1857	1560	1912	1828	1750	1902	1841	1808	1918
R3, 2009 (OC, 15)	924	362	974	932	859	976	654	628	676	919	911	976
R4, 2013 (OC, 19)	858	NA	905	921	NA	951	580	NA	635	778	NA	882

III.A *Child Time-Use*

The key feature that makes Young Lives uniquely suitable for our analysis is the data on the time-use of all children in the household at each wave, starting with round 2, when the Younger Cohort were 5 and Older Cohort were 12. We do not have detailed time diary data, but we can distinguish between time spent on eight types of activities on a "typical day"[17] in the last week: (1) sleep; (2) caring for others in the household (e.g., younger children or ill household members); (3) domestic tasks (e.g., fetching water, firewood, cleaning, cooking, washing); (4) tasks on the family farm, cattle herding, other family business; (5) paid work or activities outside of the household for someone not in the household; (6) school (including travel); (7) studying outside of school (at home, extra tuition); (8) play time/general leisure.

[17] A typical day is defined as a weekday in which the child undertook usual weekday activities. If the interview took place during a festival, respondents were asked to recall time-allocation prior to the festival period.

For our analysis, we combine categories (2) and (3) into a single category, and we will refer to the sum of hours spent on categories 2 and 3 as "domestic chores." Similarly, we combine categories (4) and (5) into a single category, and we will refer to the sum of hours spent on categories 4 and 5 as "economic activities."[18]

Table 3: Child Time Allocation – Mean Hours of each Activity on a "typical day"

	Sleeping	Domestic chores*	Economic activities*	Domestic chores + Economic activities	At school	Studying at home	Playing/Leisure
Ethiopia							
Age 8	9.63	2.47	1.47	3.94	5.07	1.05	4.31
Age 12	9.25	2.43	1.63	4.06	5.63	1.47	3.59
Age 15	8.66	3.16	1.70	4.86	5.62	1.88	2.98
Age 19	8.48	2.69	3.29	5.98	3.71	1.66	4.17
India							
Age 8	9.15	0.56	0.03	0.58	7.69	1.84	4.74
Age 12	8.95	0.97	0.18	1.15	8.06	1.85	3.99
Age 15	8.27	1.68	1.47	3.15	6.49	2.04	4.04
Age 19	8.24	2.59	3.17	5.76	3.78	1.21	5.02
Peru							
Age 8	9.97	1.37	0.25	1.63	6.21	1.94	4.25
Age 12	9.51	1.87	0.51	2.37	6.24	2.07	3.80
Age 15	9.16	2.18	1.06	3.24	6.10	2.16	3.34
Age 19	8.47	2.66	3.90	6.56	3.70	1.53	3.74
Vietnam							
Age 8	9.72	0.80	0.10	0.89	4.96	2.84	5.59
Age 12	8.97	1.36	0.42	1.78	5.47	2.70	5.08
Age 15	8.69	1.62	1.46	3.07	4.24	3.05	4.95
Age 19	8.28	1.82	4.14	5.97	2.82	1.25	5.68

*Notes: Domestic chores include caring for others in the household (e.g., younger children or ill household members) and domestic tasks (such as fetching water, firewood, cleaning, cooking and washing); Economic activities include tasks on the family farm, cattle herding, other family business (not just farming), as well as paid work or activities outside of the household for someone not in the household.

Summary statistics for time allocation for the four countries, two cohorts of children and 3 waves of data are presented in Table 3. Note that child work at ages 8 and 12 is most prevalent in Ethiopia, with average hours per day on economic activities plus domestic chores at about 4 hours. Peru is next with combined hours of 1.63 per day at age 8 and 2.37 at age 12. Combined work time at young ages is lowest in India with 0.58 hours per day at age 8 and 1.15 hours at age 12. Combined school and study time at ages 8 and 12 follows the reverse pattern. It is highest in India (about 9 ½ hours per day), followed by Vietnam (about 9 hours per day), Peru (about 8 ¼ hours per day), and finally Ethiopia (about 6 to 7 hours per

[18] (Edmonds 2009) notes there is considerable ambiguity in the classification of different child work activities and across studies. We do not aim to endorse one set of definitions versus another – but rather to be clear about how we group activities, and about the terms we use throughout the rest of the paper to reflect that grouping.

day). A very interesting difference is that school time in Vietnam is actually lower than in any other country, but home study time is much higher (about 2 ¾ hours per day).

Combined work time increases sharply in India, Peru and Vietnam at ages 15 and 19, catching up to the levels in Ethiopia. Thus, at age 19 total work time is roughly equalized across the four countries (ranging from 5 ¾ hours in India to 6 ½ hours in Peru). School time in all four countries falls sharply at age 19, and it tends to converge across countries (ranging from 2.8 hours in Vietnam to 3.8 in India).

In all four countries, children at ages 8, 12 and 15 spend (on average) more time on domestic chores than economic activities, typically by a wide margin. But economic activities become the more common type of work at age 19. The shift is especially large in Vietnam, where average hours spent on economic activities at age 19 jumps to more than double the level of hours spent on domestic chores.

Leisure time fluctuates between 3 and 6 hours per day across the four countries and age groups. It is consistently highest in Vietnam (5 to 6 hours per day depending on age), and next highest in India (4 to 5 hours per day), while in Peru it fluctuates around 3 1/3 to 4 ¼ hours per day, and in Ethiopia it fluctuates around 3 to 4 1/3 hours. In all four countries, sleep time is 9 to 10 hours at age 8 and drops to about 8 1/3 hours at age 19.

Table 4 breaks down work time along the extensive and intensive margins. It reports the proportion of children (by age/country) engaged in any work (economic or domestic), as well as average hours conditional on work. At age 8, the fraction of children who do some form of work on a typical day ranges from a low of 38% in India to 54% in Vietnam to 76% in Peru to a high of 93% in Ethiopia. By age 12 the large majority of children do some work in every country. Domestic chores are more common than economic activities in all countries and at all ages. At age 8, the fraction of children engaged in economic activities ranges from the low single digits in India and Vietnam to 16% in Peru and 40% in Ethiopia. But by age 15 about a ¼ to ½ of children are engaged in economic activities.

Conditional on working, children aged 8-12 work about 4 hours per day in Ethiopia, and roughly 2 hours per day in the other three countries. At age 15 mean hours conditional on work range from a low of 3 1/3 per day in Vietnam to nearly 5 in Ethiopia. Ethiopia stands out as the country with both the highest prevalence of child work and the highest hours conditional on work. Over 90% of children in Ethiopia are involved in some form of work at all ages, and mean hours (conditional on working) is at least 4 ¼ per day, even at age 8.

While economic activities are less common than chores, a fairly common pattern across countries/ages is that, conditional on engaging in each activity, economic activities tend to take up more hours than domestic chores. Furthermore, while less than half of children who do domestic chores also engage in economic activities, the great majority of the children who are involved in economic activities also do chores (consistent with MICS data as reported by Edmonds et al, 2007).

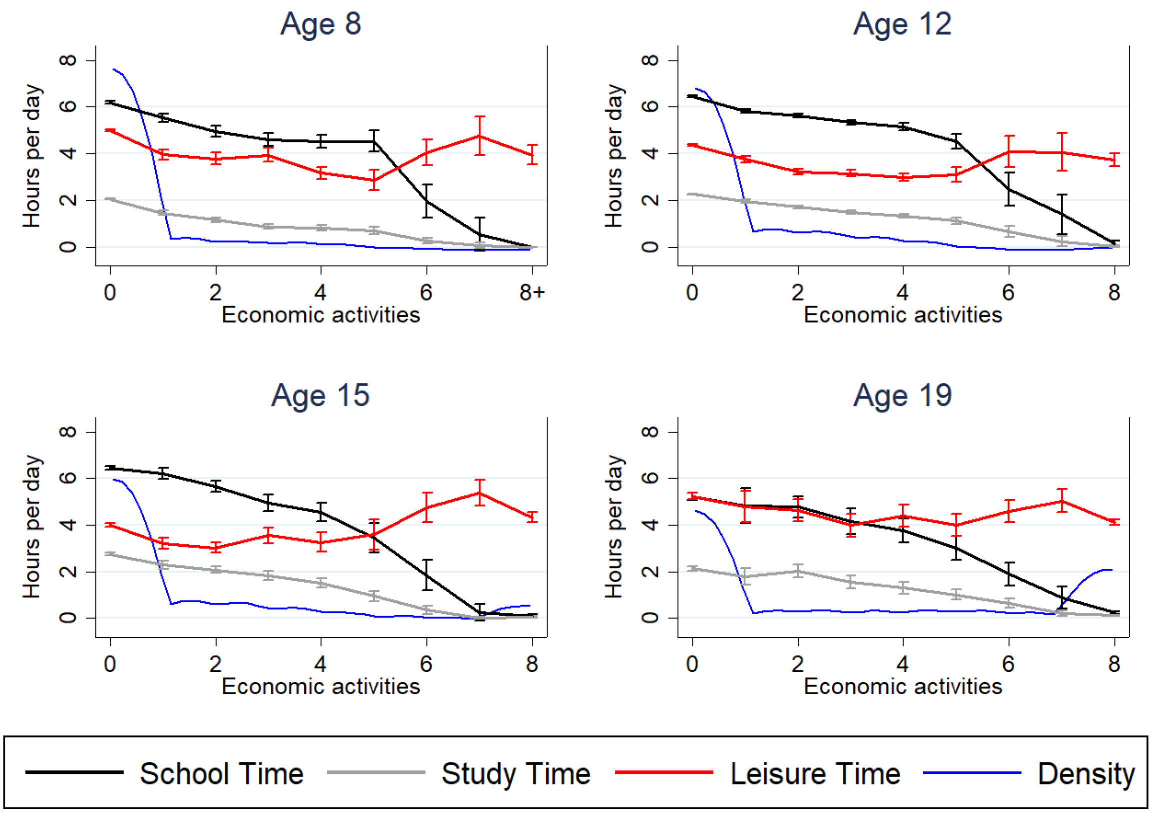
Table 4: Prevalence and intensity of working by age and country

	Work Any	Domestic chores (any)	Economic activities (any)	Domestic chores *and* Economic activities	Domestic chores hours (if any)	Economic activities (if any)	Total (if any chores or economic activities)	N
Ethiopia								
Age 8	0.93	0.85	0.40	0.33	2.89	3.67	4.25	1257
Age 12	0.94	0.85	0.48	0.39	2.87	3.41	4.34	1821
Age 15	0.99	0.92	0.45	0.38	3.45	3.77	4.92	941
Age 19	0.91	0.72	0.52	0.34	3.71	6.27	6.58	858
India								
Age 8	0.38	0.38	0.01	0.01	1.48	2.39	1.53	1683
Age 12	0.69	0.68	0.04	0.03	1.43	4.37	1.65	1860
Age 15	0.77	0.72	0.23	0.19	2.33	6.29	4.12	941
Age 19	0.88	0.82	0.41	0.36	3.14	7.67	6.57	921
Peru								
Age 8	0.76	0.75	0.16	0.14	1.84	1.57	2.12	1851
Age 12	0.90	0.88	0.25	0.23	2.12	2.02	2.65	1838
Age 15	0.91	0.89	0.32	0.30	2.46	3.28	3.56	656
Age 19	0.91	0.74	0.52	0.35	3.61	7.44	7.21	580
Vietnam								
Age 8	0.54	0.52	0.04	0.03	1.53	2.30	1.66	1706
Age 12	0.77	0.74	0.21	0.18	1.84	2.06	2.32	1847
Age 15	0.93	0.87	0.33	0.26	1.86	4.45	3.29	920
Age 19	0.93	0.80	0.57	0.44	2.28	7.33	6.44	778

Figure 1 presents descriptive statistics on how school, study and leisure time vary with total hours spent on economic activities, combining data from all four countries. The horizontal axis reports hours spent on economic activities, which vary from 0 to 8+ hours per day. The blue line shows the density of children at each level of work hours. For example, the upper left panel shows results for 8 year olds. As we see, both school and study time tend to fall steadily as hours of economic activity increase (e.g., at 4 hours of work, both school and leisure time are reduced by about 2 hours). Interestingly, however, average school time falls sharply and leisure *increases* when economic activities reach 6+ hours per day. The implication is that it is very difficult to stay in school while working 6+ hours, so we see a discrete break whereby many children cease school entirely, while leisure is increased.[19]

[19] This pattern would be predicted by a simple model with fixed time or monetary costs of school attendance, which render low levels of school hours inefficient. Study time appears to fall more steadily with work hours.

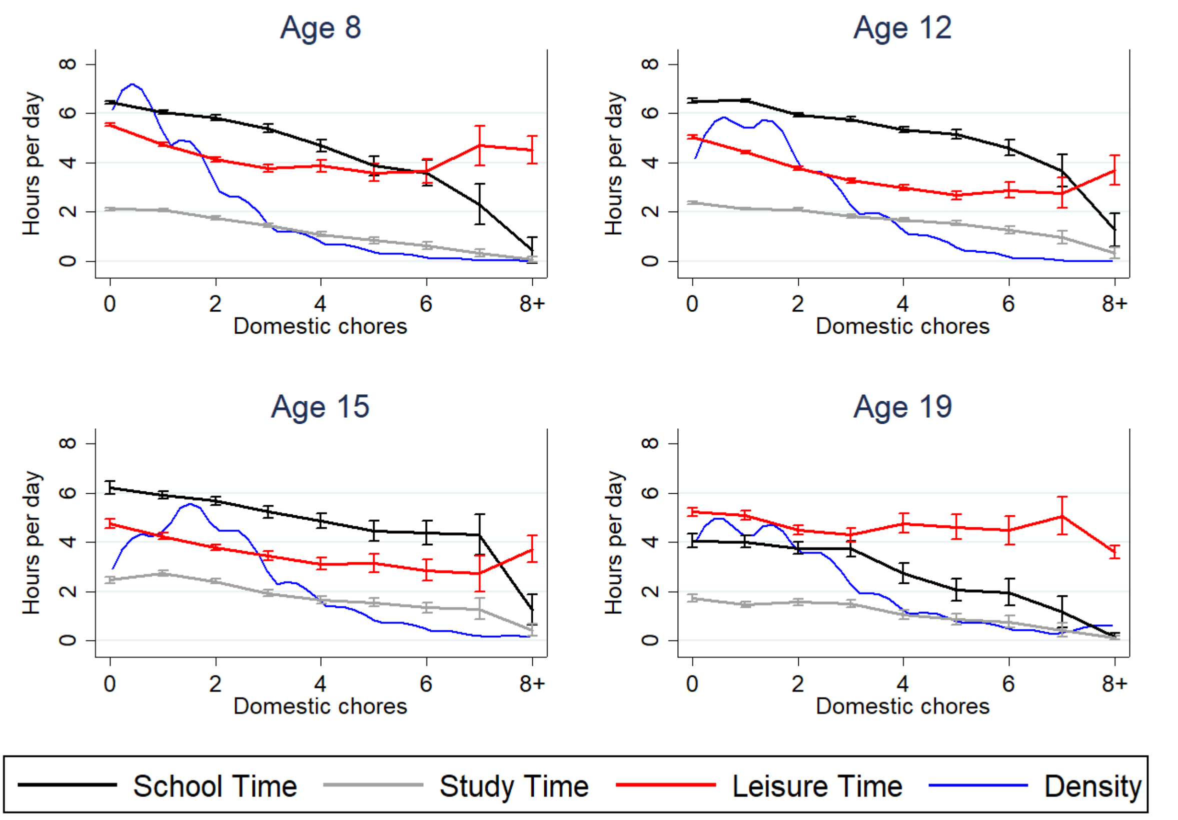
Figure 1: Evolution of School and Leisure Hours with Economic Activities



Of course, very few 8 year olds spend 6+ hours per day in economic activities, but we see the same basic pattern at ages 12, 15 and 19. That is, school hours fall gradually at low levels of economic activity and more sharply at higher levels. In contrast, leisure falls at low levels of work hours, but leisure increases once work hours reach high levels. Thus, we see an important non-linearity in the relationship between work hours and school hours: low to moderate work hours crowd out both school and leisure time. But at sufficiently high levels of work, it is school that is crowded out while leisure actually increases. We thus hypothesize that child labour will have a much more negative impact on child cognitive development once work hours reach the level that sharply impacts on school hours.

Figure 2 presents descriptive statistics on how school, study and leisure time vary with total hours spent on domestic chores. Superficially, the patterns in Figures 1 and 2 look quite similar. But a careful comparison reveals that (i) at ages 8 and 12, school time falls more slowly with domestic chores than with economic activities, and (ii) at ages 8, 12 and 15, a higher level of domestic chores can be sustained before we see a sharp drop in school hours and an increase in leisure. For example, at age 12, school hours drop sharply and leisure increases when economic activities go from 5 to 6+ per day (see Figure 1). But in Figure 2 such a pattern does not emerge until domestic chores go from 7 to 8+ hours per day. We thus hypothesize that, in models that fail to account for the complete time budget, domestic chores will appear to have a less negative impact on child cognitive development than economic activities, simply because domestic chores seem to do less to crowd out school time.

Figure 2: Evolution of School and Leisure Hours with Domestic Chores



<u>III.B. *Child Skill Assessments*</u>

A distinguishing feature of the Young Lives data is the richness of the skill/ability measures that are available. The longest and most comparable time-series are available for assessments of children's mathematics and verbal skills. Math assessments were administered to both cohorts in 2006, 2009 and 2013; these are, therefore, available for the YC at ages 5, 8 and 12 and for the older cohort at ages 12, 15 and 19.

Verbal skills are captured through a widely used receptive vocabulary test, known as the Peabody Picture Vocabulary Test (PPVT).[20] A number of studies have found a strong positive correlation between PPVT and some commonly-used intelligence measures, such as the Wechsler and the McCarthy Scales (Campbell 1998; Campbell, Bell, and Keith 2001;

[20] The PPVT-III (Dunn, Dunn, and Service 1997) was adapted for administration in Vietnam, Ethiopia and India. In Peru, the Spanish version PPVT-R adapted for Latin America was used (Dunn et al. 1986). See technical notes on Young Lives adaptation procedures on www.younglives.org.uk.

Gray et al. 1999). The PPVT was administered to both cohorts in 2006 and 2009 and to YC only in 2013. It is, therefore, available for the younger cohort at ages 5, 8 and 12 and for the older cohort at ages 12 and 15.

In the PPVT, the child is asked to select the picture that best represents the meaning of a stimulus word presented orally by the examiner. Test items were arranged in order of increasing difficulty and only the items within the critical range of the specific child were administered to each child, selected by the interviewer. The test score is calculated as the difference between the ceiling item (e.g. word number 78) and the total number of errors.[21]

Age 5 math skills were assessed using the Cognitive Development Assessment (CDA), which is an instrument developed by the International Association for the Evaluation of Educational Achievement (IEA) to assess the cognitive development of young children (Montie, Xiang, and Schweinhart 2006). The sub-scale administered to the Young Lives children YC cohort in round 2 tests children's understanding of concepts such as few, most, half, many, equal, and pairs with statements such as "Point to the plate that has few cupcakes." The sub-scale has 15 items and each correct answer is scored 1 point so that the minimum number of points a child can get is 0 and the maximum is 15.

After age 5, math skills were assessed using self-administered paper-based tests especially designed by the education experts on the Young Lives research team. In order to cover the wide range of math proficiency across countries and ages, the test contains items of highly varied levels of difficulty.

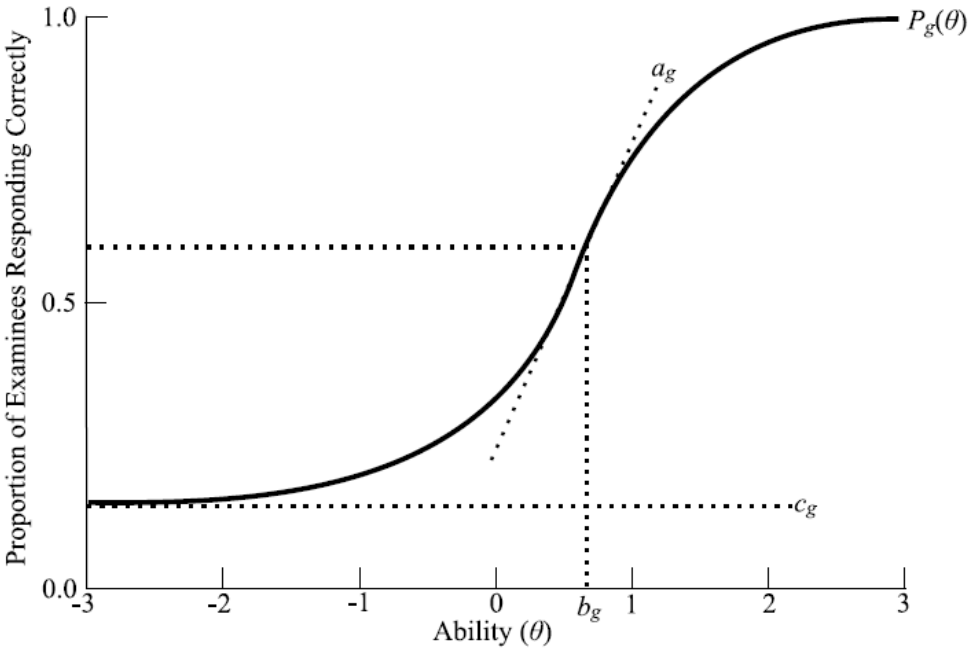
III.C. *Item Response Theory*

A major challenge we face in using data that varies by country and age to estimate a child cognitive ability production function is that we must develop cognitive ability measures that are comparable across countries and ages. In order to achieve this we use Item Response Theory (IRT) to model the math and verbal test score questions, and then derive consistent measures of latent math and verbal ability that are comparable across age and countries. IRT has a long history in education and psychometrics (e.g., van der Linden and Hambleton 1997). For instance it underlies the construction of well-known tests like the SAT, and it has been applied to generate internationally comparable scores for tests such as PISA and TIMMS, and across state/cohort comparable scores for the NAEP and ECLS. But with few exceptions (e.g., Das and Zajonc 2010) IRT remains little used by economists.

[21] Official PPVT test manuals include tables for standardisation of the raw scores. But these procedures were not followed in the Young Lives study as the standardisation samples have different characteristics from the Young Lives samples.

In the IRT model, each item (multiple choice question) on a test is assumed to be characterized by an "Item Characteristic Curve" (ICC) that maps person *i*'s latent ability θ_i into the probability he/she answers the question correctly. As shown in Figure 3, we assume a standard three parameter ICC, where the parameters (a, b, c) map ability into responses:

Figure 3: Three Parameter Item Characteristic Curve (3PL Model)



The parameter c_g in Figure 3 is the probability a person can guess the correct answer to item g, given he/she has essentially no ability to actually determine the correct answer. Parameter b_g is known as the "difficulty" of item g. It measures the ability level at which the probability of a correct answer passes the half-way point between the probability of a correct guess and 100%.[22] The parameter a_g is known as the "discriminating" power of item g. For instance, if a_g is large, it implies the probability a person will answer item g correctly rises quickly as their ability passes the level $\theta_i = b_g$. Thus, the question is good at discriminating between people who are just above and below that ability level. The response function $P_g(\theta|a_g,b_g,c_g)$ can be written as $P_g(\theta|a,b,c) = c_g + (1 - c_g)F[\mathrm{a}_g(\theta_i - b_g)]$, where $F(\cdot)$ is the standard normal cumulative distribution function (CDF).

Given a sample of responses of a set of individuals to a set of test questions (with sufficient overlap of questions across people), one can use maximum likelihood to jointly

[22] In a two-parameter ICC, where the possibility of correct guessing is ignored (i.e., c_g=0), we have that $\theta_i = b_g$ is simply the ability level at which the probability of giving a correct answer to question g is exactly 50%.

estimate the characteristics (a, b, c) of all items and the ability levels of all individuals.[23] Given an existing test with known ICCs (i.e., known values of (a, b, c)), one can estimate an individual's ability level as the value of ability θ_i that maximizes the posterior probability of the person's entire set of test responses.

There are three main advantages of using IRT to generate ability scores: (a) it offers a less arbitrary way to construct skill measures than simply summing up correct responses, (b) it offers a way of running diagnostics on the performance of specific test items for cross-cultural comparisons (see below), and (c) conditional on having partial overlap of questions across time and/or contexts, it allows scores to be linked on a common metric making it possible to construct comparable measures of achievement across age/cohorts/countries.

Assessments administered within Young Lives were designed to include partial overlap of questions over time and (for math) across countries in order to allow linking of scores to a common metric. Utilising these common items and applying IRT, the tests can be linked in the following ways:

- Math ability tests at age 5 can be linked across countries
- Math ability tests at ages 8-19 can be linked across countries and across ages
- Verbal ability tests (within-language) can be linked across ages 5 to 15

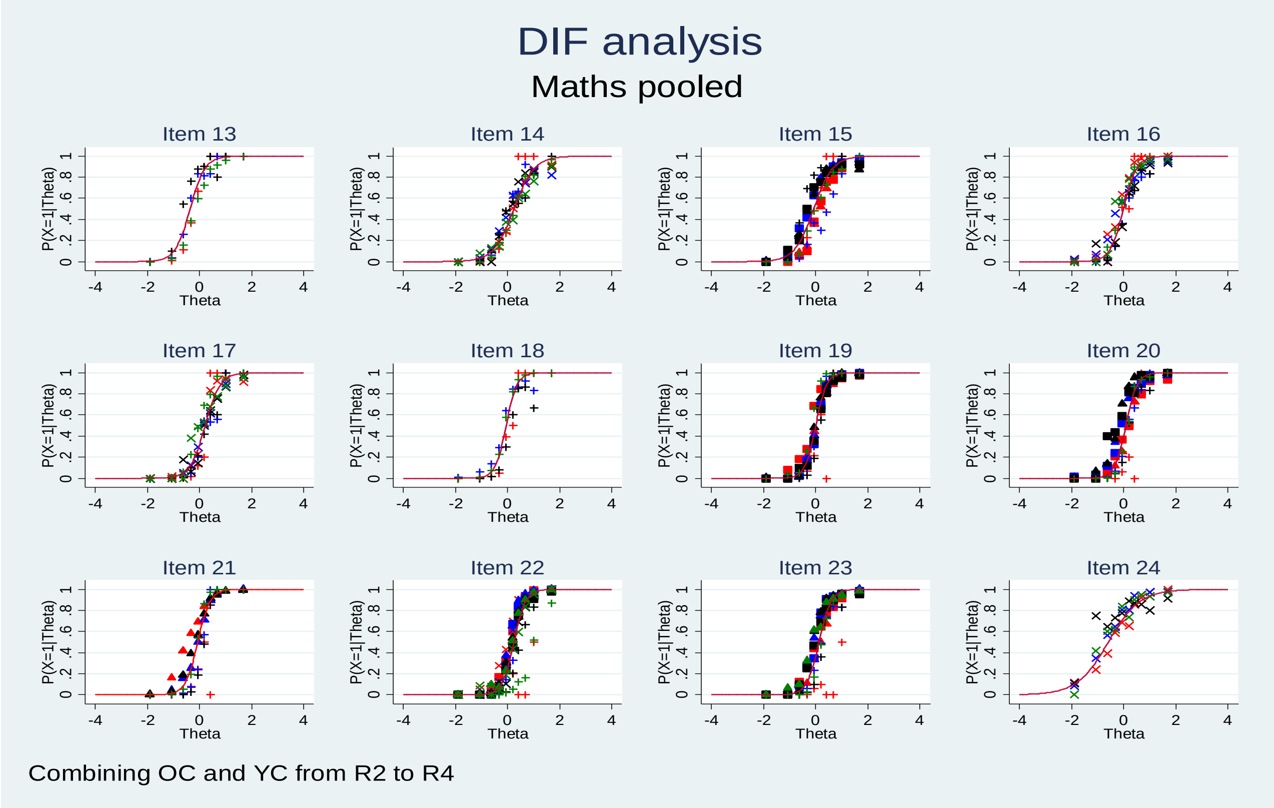
The IRT-linked ability scores are normalized to have a mean of zero and a standard deviation of one in a base age group. Specifically, math scores are normalized with reference to the distribution of test scores of 8-year olds pooled across countries, and receptive vocabulary (PPVT) scores are normalized with reference to the 5-year old age group within language.

Figure 4 presents a test of pooling math test results across the four countries, two cohorts and three rounds and four ages. It plots known item characteristic curves for twelve items that are common across all (or most) tests taken by all children. If the ICC curves are invariant to context, then the probability of a correct response should vary with measured ability in the same way across countries/cohorts/rounds (ages). Figure 4 shows that ability quantiles line up along the ICC curves quite accurately in nearly all instances. The results for the verbal scores are similarly encouraging.[24]

[23] Formally, IRT relies on three key assumptions: (i) uni-dimensionality; i.e., a single latent person-specific trait determines individual performance on the test, (ii) no differential item functioning; i.e., the item characteristics (a, b, c) are person-invariant, and (iii) conditional local independence; i.e., item responses are independent across persons and, conditional on ability, across questions for the same person. Under these assumptions, one can consistently estimate both item characteristics and individual ability levels, given data on test results of a sample of subjects.

[24] We thank the Young Lives team and affiliated researchers, especially Santiago Cueto, Juan Leon, Caine Rolleston and Abhijeet Singh for their work on designing the tests as well as constructing the linked scores.

Figure 4: Example of Testing IRT Assumptions



Notes: Ethiopia – Red, India – Blue, Peru – Black, Vietnam – Green
Young Cohort: (R3) +, (R4) ▲, Older Cohort: (R2) ●, (R3) **x**, (R4) ■
For each item/country/cohort/round we plot the empirical frequency of a correct answer at deciles of ability. If pooling is acceptable then all ability points should lie on the ICC for each item. The empirical probabilities corresponding to each decile are often not clearly visible, as they often lie essentially on top of one another.

III.D. *Descriptive Statistics*

Appendix Table 1 reports summary statistics for the data collected in the second round in 2006 when the YC were 5 and the OC were 12. This is the earliest round of data that we use in the analysis. As we see in the table, the sample is evenly split between boys and girls across the four countries and cohorts. On average parental education is much lower in Ethiopia and India than Peru and Vietnam. For example, while around a half (44-58%) of the mothers in both cohorts in Ethiopia and India have no schooling, this is the case for only about a tenth of the mothers in Peru and Vietnam.

Household size and composition in Ethiopia look somewhat different from the other three countries: Ethiopian children tend to live in larger households, with a more brothers and sisters but fewer grandparents. The biggest age gap between mothers and fathers is also found in Ethiopia, where on average mothers are ten years younger than fathers.

The majority of households in Ethiopia, India and Vietnam are in rural areas. The Peru sample, however, is much more urban, with around 70% of households located in urban or peri-urban areas. In all countries, the children are on average more than a standard deviation shorter than the average child of their age in a healthy population (according to the WHO).[25] In Peru by far the biggest group (over 90%) identify themselves as "Mestizo" and in Vietnam over 80% of the sample are from the majority Kinh group, while in Ethiopia and India there is more variation in ethnicity.[26]

IV. Estimation

We can define the production function for *measured* skill Y of individual i at age a as:

$$Y_{ia} = F_a(\boldsymbol{X}_i(a), \boldsymbol{U}_i(a), \mu_{i0}, m_{ia}) \quad (1)$$

Where $\boldsymbol{X}_i(a) = (\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{ia})$ and $\boldsymbol{U}_i(a) = (\boldsymbol{u}_{i1}, \dots, \boldsymbol{u}_{ia})$ are vectors that contain the full history of observed and unobserved inputs, respectively, μ_{i0} is child's innate "ability" and m_{ia} captures transitory measurement error in the skill test.

Assuming that the production function is linear, the above can be written as:

$$Y_{ia} = \sum_{k=0}^{a} \boldsymbol{x}'_{i,a-k}\boldsymbol{\beta}_{a-k} + \sum_{k=0}^{a} \boldsymbol{u}'_{i,a-k}\boldsymbol{\gamma}_{a-k} + \rho_a\mu_{i0} + m_{ia} \quad (2)$$

As is well documented in the skill formation literature, the main challenge of estimating this model is omitted variables. We do not observe μ_{i0} or $\boldsymbol{U}_i(a)$, as no dataset can contain a complete history of relevant inputs. An endogeneity problem arises if these omitted inputs are correlated with the observed inputs. Typically, all the lags of $\boldsymbol{x}_{ia}$ are not observed either.

We address these problems by estimating a value added (VA) specification in which a lagged skill measure $Y_{i,a-1}$ is included as a proxy for the unobserved μ_{i0} and $\boldsymbol{U}_i(a)$ and the missing lagged $x's$. The standard VA model has the form:

$$Y_{ia} = \rho Y_{i,a-1} + \boldsymbol{x}'_{ia}\boldsymbol{\beta} + \varepsilon_{ia} \quad (3)$$

We now clarify assumptions under which this VA approach gives consistent estimates of the effects of current inputs. This depends on the properties of ε_{ia}. Our discussion closely follows Todd and Wolpin (2007). First, using (2) to substitute for $Y_{i,a-1}$ in (3) we obtain:

[25] Height for age z-scores were calculated using the 2006 WHO international standard for children younger than age 5 (WHO, 2006), and the 2007 WHO international standard for children older than 5.
[26] In three countries there is a dominant religion reported by over 70 percent of the sample: Christian Orthodox in Ethiopia, Hindu in India, and Catholic in Peru. In Vietnam the majority report no religious affiliation.

$$Y_{ia} = \rho\left[\sum_{k=1}^{a} \boldsymbol{x}'_{i,a-k}\boldsymbol{\beta}_{a-k} + \sum_{k=1}^{a} \boldsymbol{u}'_{i,a-k}\boldsymbol{\gamma}_{a-k} + \rho_{a-1}\mu_{i0} + m_{i,a-1}\right] + \boldsymbol{x}'_{ia}\boldsymbol{\beta} + \varepsilon_{ia} \quad (4)$$

We see that (3) and (4) are equivalent for all *a iff* we define $\varepsilon_{ia} = \boldsymbol{u}'_{i,a}\boldsymbol{\gamma}_a + m_{ia} - \rho m_{i,a-1}$ and assume $\rho_a = \rho^a \; \forall \; a$, and $\boldsymbol{\beta}_{a-k} = \rho^k \boldsymbol{\beta} \; \forall \; a$, and $\boldsymbol{\gamma}_{a-k} = \rho^k \boldsymbol{\gamma} \; \forall \; a$. Then (4) becomes:

$$Y_{ia} = \boldsymbol{\beta}\boldsymbol{x}_{ia} + \rho\left[\sum_{k=1}^{a} \boldsymbol{x}'_{i,a-k}\boldsymbol{\beta}\rho^{k-1} + \sum_{k=1}^{a} \boldsymbol{u}'_{i,a-k}\boldsymbol{\gamma}\rho^{k-1} + \rho^{a-1}\mu_{i0} + m_{i,a-1}\right] + \varepsilon_{ia} \quad (5)$$

Equation (5) clarifies how (3) assumes the effects of lagged inputs $\boldsymbol{x}_{i,a-k}$ and the initial skill endowment μ_{i0} on the current outcome (test score) Y_{ia} all depreciate at the same rate ρ.

Thus, as discussed by Todd and Wolpin (2007), the VA model assumes: (i) current effects of inputs are age invariant ($\boldsymbol{\beta}_a = \boldsymbol{\beta} \; \forall \; a$, $\boldsymbol{\gamma}_a = \boldsymbol{\gamma} \; \forall \; a$), and (ii) the effects of observed ($\boldsymbol{x}$) and unobserved ($\boldsymbol{u}$) inputs and unobserved ability (μ) all depreciate at the same rate (ρ). Furthermore, as the error term in (3) is $\varepsilon_{ia} = \boldsymbol{u}'_{i,a}\boldsymbol{\gamma}_a + m_{ia} - \rho m_{i,a-1}$, it includes both *unobserved* current inputs ($\boldsymbol{u}_{ia}$) and measurement error in the lagged test score ($m_{i,a-1}$), creating potential endogeneity problems. We now discuss these issues in turn:

Following Todd and Wolpin (2007), we can relax the common depreciation rate of inputs assumption by estimating an "extended" version of the VA Model that includes lagged inputs $x_{i,t-1}$ directly in the outcome equation:

$$Y_{ia} = \rho Y_{i,a-1} + \boldsymbol{x}'_{ia}\boldsymbol{\beta} + \boldsymbol{x}'_{i,a-1}\boldsymbol{\pi} + \varepsilon_{ia} \quad (6)$$

If we substitute out for $Y_{i,a-1} = \rho Y_{i,a-2} + \boldsymbol{x}'_{i,a-1}\boldsymbol{\beta} + \boldsymbol{x}'_{i,a-2}\boldsymbol{\pi} + \varepsilon_{ia-1}$ we obtain:

$$Y_{ia} = \boldsymbol{x}'_{ia}\boldsymbol{\beta} + \boldsymbol{x}'_{i,a-1}(\rho\boldsymbol{\beta} + \boldsymbol{\pi}) + \left\{\rho^2 Y_{i,a-1} + \boldsymbol{x}'_{i,a-2}\boldsymbol{\pi}\rho\right\} + \left[\varepsilon_{ia} + \rho\varepsilon_{i,a-1}\right] \quad (7)$$

where π allows the depreciation rate to differ from ρ. This is not only a more flexible specification, but also allows us to test whether the effects of inputs on outcomes deteriorate more or less slowly than predicted by a standard VA model.[27]

In addition, we further extend the VA model by also allowing for age-varying coefficients (ρ_a, β_a, π_a). Thus, the model we estimate on children at age *a* has the form:

$$Y_{ia} = \rho_a Y_{i,a-1} + \boldsymbol{x}'_{ia}\boldsymbol{\beta}_{\boldsymbol{a}} + \boldsymbol{x}'_{i,a-1}\boldsymbol{\pi}_{\boldsymbol{a}} + \varepsilon_{ia} \quad (8)$$

If we substitute out for $Y_{i,a-1}$ we obtain equation (9), which illustrates how the effects of

[27] E.g., if $\beta > 0$ and $\pi > 0$ the (positive) effect of *x* depreciates more slowly than the rate ρ implied by the VA model. If $\beta < 0$ and $\pi > 0$ the (negative) effect of *x* depreciates more quickly than implied by the VA model.

inputs and the depreciation rate are *both* allowed to vary by age:

$$Y_{ia} = \boldsymbol{x}'_{ia}\boldsymbol{\beta}_a + \boldsymbol{x}'_{i,a-1}(\rho_a\boldsymbol{\beta}_{a-1} + \boldsymbol{\pi}_a) + \{\rho_a\rho_{a-1}Y_{i,a-2} + \boldsymbol{x}'_{i,a-2}\boldsymbol{\pi}_{a-1}\rho_a\} + [\varepsilon_{ia} + \rho_a\varepsilon_{i,a-1}] \quad (9)$$

Interestingly, we are unaware of prior applications of the fairly flexible model in (8).

Next, we consider the problem of measurement error in the lagged test score. As we noted earlier, the error term ε_{ia} in equation (3) includes the term $-\rho m_{i,a-1}$. Hence it is likely that ε_{ia} will be negatively correlated with the lagged test score, biasing the estimate of ρ downward, and biasing $\widehat{\boldsymbol{\beta}}$ in an ambiguous direction. A practical approach to this problem is simply to make the test as accurate as possible (by using IRT methods), and to include in the model other good proxies for child ability, such as parents' education. But we will also report results where we instrument for the lagged test score using another test (see Section V.F).

Finally, as noted earlier, ε_{ia} will in general contain unmeasured current period inputs, $\boldsymbol{u}_{ia}$, as the lagged test score only proxies for lagged inputs. If $\boldsymbol{u}_{ia}$ is correlated with measured current inputs ($\boldsymbol{x}_{ia}$) we have an endogeneity problem. Using IV to deal with this problem is difficult, as it is hard to imagine instruments that shift *observed* inputs but leave *unobserved* inputs unaffected.[28] Obviously, endogeneity is less of a concern if we have extensive controls for non-time inputs and family resources. This brings us to our empirical specification:

Our estimating equation, an extended version of equation (8), can be written:

$$Y_{ia} = \rho_a Y_{i,a-1} + \boldsymbol{x}'_{ia}\boldsymbol{\beta}_a + \boldsymbol{x}'_{i,a-1}\boldsymbol{\pi}_a + \boldsymbol{b}'_{ia}\boldsymbol{\delta}_{ia} + \varepsilon_{ia} \quad (10)$$

As in (8), $\boldsymbol{x}_{ia}$ is a vector of time-varying inputs at age a, and $\boldsymbol{x}_{i,a-1}$ is its lag. The vector $\boldsymbol{b}_{ia}$, which we have supressed until this point for notational convenience, contains controls for background characteristics of the child (own, parental and household) which, in addition to $Y_{i,a-1}$, serve as additional proxies to help capture omitted inputs and unobserved ability (see details below). We estimate equation (10) separately for children of each age in each country.

As we noted earlier, the value added approach has been advocated by Orazem and Gunnarsson (2004), and recent studies in the education literature offer encouraging evidence on the effectiveness of the lagged test score as a control for unobserved heterogeneity (Guarino, Reckase, and Wooldridge 2014; Muralidharan and Venkatesh 2013; Deming et al. 2014; Angrist, Pathak, and Walters 2013). Furthermore, one may argue that the VA model is

[28] For example, suppose the nutrition input into child development is unmeasured. A positive shock to the wage of child labour may cause an increase in child work that is exogenous in the sense it is uncorrelated with ability, but it may also improve the nutrition input because household income increases. In this example we may underestimate the *ceteris paribus* negative effect of child work because it coincides with improved nutrition.

more intuitive than a conventional fixed effects model in the child development context, as it seems implausible that the positive influence of the initial skill endowment μ_{i0} does not depreciate in the absence of subsequent investments in the child (whereas the conventional fixed effects model assumes that μ_{i0} has a time invariant effect).

In our specification of (10) the time-varying inputs $\boldsymbol{x}_{ia}$ correspond exactly to the vector of child time use variables: time spent working in economic activities and domestic chores, time in school, time studying at home, playing/leisure and sleeping. The background characteristics $\boldsymbol{b}_{ia}$ include those of the child (age, sex, height-for-age z-score, religion and ethnicity), the parents (father's and mother's age and education, whether parents live in the household), and the household (whether in an urban community, a wealth index,[29] household member composition).[30] The elements of $\boldsymbol{b}_{ia}$ either vary deterministically over time (age), change little over time, or are time-invariant. Thus, we do not include $\boldsymbol{b}_{i,a-1}$ in (10). The full set of background characteristics included in the main models are listed in Appendix Table 1.

V. Results

We estimate equation (10) separately for children of each age in each country. Recall that we can estimate the math score equation for ages 8, 12, 15 and 19, and the verbal score equation for ages 8, 12 and 15. Thus, we estimate a total of 28 regressions. Given that each regression contains a large number of time inputs and control variables, it is not feasible to present the complete set of parameter estimates.[31] Instead, we graphically summarize the estimates of the main coefficients of interest. These are the coefficients on the hours of child work, separated into domestic chores and economic activities.

Figure 5 plots the child work coefficients from child cognitive ability production functions with math scores as the dependent variable. The left panel plots the coefficients on domestic chores, while right panel plots those on economic activities. Note there are 16 estimates in each figure, as we have 4 countries and 4 ages.[32] The figures also report 95% confidence intervals around the point estimates.

[29] We use the wealth index constructed and publicly archived by the Young Lives team. See Note for Appendix Table 1 for details of how it is constructed.

[30] Given the volatility of households' current income, education and wealth are likely to be better proxies than income for goods inputs (assuming goods inputs are largely determined by permanent income).

[31] The complete set of estimates is available by request.

[32] Each coefficient in a plot comes from an age/country specific regression as indicated on the x-axis. For example in Figure 5 Panel A, in left-hand side figure labelled "Chores," the first coefficient, which is labelled "ET 8," is the effect of chore/care time on math sores estimated for a sample of Ethiopian children, by regressing the age 8 math test score on the time-allocation vector at age 8, the lagged time-allocation vector at age 5, the lagged math test score at age 5, and vector of background characteristics at age 8.

To identify the time input coefficients we must choose an omitted time category.[33] All parameters estimates can then be interpreted as the effect of a particular type of child work *relative* to the omitted activity. While the choice of omitted category has no substantive importance, it does have practical importance for the proper interpretation of the estimates.

In Figure 5 Panel A the omitted time use category is leisure. Thus, the coefficients in the left figure labelled "chores" should be interpreted as the effect of increasing time spent on domestic chores by one hour while reducing leisure by one hour, and the coefficients in the right figure labelled "economic activities" should be interpreted as the effect of increasing time spent on economic activities by one hour while reducing leisure by one hour.

V. A. *Child Work and Math Development*

Figure 5 Panel A illustrates a striking result: There is no clear evidence that time spent in child work is less productive for math ability than time spent in leisure. Only one of the 16 estimates of the effect of time spent on domestic chores relative to leisure is significantly negative (5% level), and only 2 of the 16 estimates of the effect of time spent on economic activities relative to leisure are significantly negative.[34] Almost all the coefficients are also small in magnitude, and 15 of the 32 coefficients on domestic chores and economic activities are the "wrong" sign (i.e., the point estimate of the child work effect is actually positive). One of these "wrong" signs is even significant.

In sharp contrast, there is strong evidence that child work is detrimental for the development of math ability if it crowds out time at school. In Figure 5 Panel B we run the exact same regressions, except here we make school time the omitted category. Thus, the estimates in the figures should now be interpreted as the effects of substituting an hour of child work for an hour of school time.

First, we find that 11 of the 16 estimates of the effect of economic activities relative to school are significantly negative at the 1% level, while 2 others are marginally significantly negative, and three are essentially zero. The median point estimate implies an extra hour of economic activities reduces math scores by 0.04 standard deviations.

To put this in some context, recall from Table 4 that the average child engaged in economic activities in Ethiopia works about 3.5 hours per day at ages 8, 12 and 15. The modal point estimate (0.04) implies such an increase in work relative to school would lower math scores at a given age by 0.14 standard deviations. This is equivalent to 16% of the

[33] As we allocate all 24 hours of the day to the possible activities, the five time-inputs are perfectly collinear.
[34] Given we are reporting 32 estimates, it would not be the least bit surprising to find 2 false positives due to type I error under the (true) null hypothesis that the effect of child work is no different from that of leisure.

average gain in math scores between age 8 and 12, and 25% of the average gain between age 15 and 19. This effect will further be magnified by the dynamics of the model.

Second, the results in Figure 5 Panel B also provide clear evidence that time spent on domestic chores instead of school is detrimental for development of math ability. We find that 7 of the 16 estimates of the effect of domestic chores relative to school are significantly negative at the 5% level, while 3 others are significantly negative at the 10% level, and all 16 point estimates are negative. The median point estimate implies that an extra hour of domestic chores reduces math scores by about 0.03 standard deviations.

It is interesting the median estimate of the negative effect of domestic chores is only slightly smaller than the median estimate for economic activities. Some prior work has suggested that economic activities have a more adverse impact (Bezerra et al. 2009; Buonomo Zabaleta 2011; Emerson et al. 2017; Gunnarsson et al. 2006; Orazem and Gunnarsson 2004). We will test this directly below.

Next, in Figure 5 Panel C, we make study time the omitted category. Comparing panels B and C, we see that both forms of child work have a more adverse effect on math scores to the extent that they substitute for study time rather than school time. This is just another way of saying that study time is more productive for child development than school time *per se* in the LMIC context. We also look at this issue directly below.

V.B. *Child Work and Verbal Development*

Our results for child work and verbal development are broadly similar to those for math. Figure 6 repeats the analysis of Figure 5 except now we look at verbal skills as measured by the PPVT test. Note that each graph in Figure 6 reports 12 coefficients, because we have 4 countries and three ages (8, 12 and 15).

Figure 6 Panel A illustrates the striking result that time spent in child work is no less productive for verbal development than time spent in leisure. Only one of the 12 estimates of the effects of domestic chore time relative to leisure is significantly negative (and only at the 10% level), and one estimate is significantly positive.[35] Only 1 of the 12 estimates of the effect of economic activities time relative to leisure is significantly negative at the 5% level. Almost all the coefficients are also small in magnitude, and 6 of the 12 coefficients on chores are the "wrong" sign (i.e., the point estimate of the child work effect is actually positive).

This is a key finding which – combined with our math results – suggests that on the whole, working children are not engaged in activities which are more detrimental to their

[35] Given we are reporting 24 estimates, it would not be the least bit surprising to find one false positive due to type I error under the (true) null hypothesis that the effect of child labour is no different from that of leisure.

cognitive development (either math or verbal) than what they do during their leisure time. It appears to hold for younger and older children and across very different contexts (in which children are likely to be undertaking different tasks as part of chores and economic activities).

In Figure 6 Panel B we run the exact same regressions, except here we make school time the omitted category. The results provide strong evidence that shifting children's time from school to economic activities is detrimental for development of verbal ability. We find that 6 of the 12 estimates of the effect of economic activities relative to school time are significantly negative (5 at the 1% level, 1 at the 5% level). All 12 point estimates are negative, and most are large in magnitude. The median point estimate implies that an extra hour of economic activities reduces verbal scores by roughly 0.05 standard deviations, which is slightly larger than what we found for math scores.

Interestingly, the results are much weaker for effects of domestic chores on verbal ability. Here only 1 of the 12 estimates of the effect of domestic chores relative to school is significantly negative at the 5% level, while 1 other is significantly negative at the 10% level. Still, all but one of the point estimates is the "right" sign (and the other is essentially zero). But the modal point estimate implies that an extra hour of chores reduces verbal scores by about 0.02 standard deviations, which is less than half the size of the economic activity effect.

Next, in Figure 6 Panel C, we make study time the omitted category. Comparing panels B and C, we see that both forms of child work have a more adverse effect on verbal scores to the extent that they substitute for study time rather than school time. Thus, similar to Figure 5 Panel C, we find evidence that study time is more productive for child development than school time *per se* in these LMIC contexts. We now examine this issue more directly.

V.C. *School vs. Study Time and Chores vs. Economic Activities*

In Sections V.A and V.B we found evidence that home study time is more productive than actual school time for child development. This result may not be surprising given the evidence of exceptionally low school quality in many LMIC's (Glewwe and Muralidharan 2016). In low school quality contexts, study time at home may be particularly important for skills accumulation.

We examine this issue directly in Figure 7, where we make study time the omitted time use category, and report the school time coefficients. Thus, the estimates can be interpreted as the effect of substituting an hour of home study time for an hour of school time. Clearly, in all four countries and at each of the four ages, time spent studying at home is at least as productive as, if not more-so than, time spent in school. Combining the math and verbal results, we see that in 5 out of 28 cases, time spent studying at home is significantly more productive than time spent at school at the 5% or 1% level, and in one case it is more productive at the 10% level. In the remaining 22 cases they are not significantly different.

This pattern is most pronounced in Ethiopia, where 3 out of 4 estimates of the effect of school time relative to home study time on math skills are negative and significant (at the 5% level). This is consistent with findings in other studies using Young Lives data which show Ethiopia and India have lower school productivity than Peru and Vietnam (Singh 2014).

Next, in Figure 8, we take a closer look at the relative impact of domestic chores vs. economic activities. Specifically, we report results from a model where domestic chore time is the omitted category, and we examine the coefficients on economic activities. The estimates can be interpreted as the effects of substituting an hour of economic activities for an hour of domestic chores (holding school time and other activities fixed). For math there is little evidence that economic activities are worse for development than chores. Only 3 of the 16 coefficients are significantly negative, and the remaining 13 are rather tightly clustered near zero. For verbal skills the point estimates tend to be larger negative values, but they are also less precisely estimated. Hence, only 1 of the 16 coefficients is significant at 1% level, with one additional one significant at the 10% level. Thus, we find no clear evidence that economic activities are worse for child development than time spent on domestic chores.

V.D. *Comparison to "Status Quo" Results*

A key feature of our paper is that we link our estimates to a specified counter-factual activity. Because we control for the full vector of activities that children spend their time on, we are able to estimate the developmental returns to time spent working *relative* to each of the other activities children may engage in (e.g., leisure, school or studying at home).

In contrast, the typical approach in most of the child labour literature is to estimate the effect of work time without controlling for other time uses. Moreover, the estimated models often include an indicator for whether a child works, rather than controlling for the actual number of work hours (see Amin, Quayes, and Rives 2006; Beegle et al. 2009; Dumas 2012; Emerson et al. 2017; Patrinos and Psacharopoulos 1997; Ravallion and Wodon 2000). Obviously our approach is more data intensive, and is only made feasible by the high quality of the Young Lives data.

To assess the usefulness of our approach, we now examine whether (and how) it changes insights gained relative to the less data demanding approach that has been typical in the child labour literature. We do this by comparing conclusions from our main results to those based on results of re-estimating all of the main specifications including only dummies for whether children spend any time in economic activities or doing domestic chores. We call analysis which controls only for child work (and not other types of time-use) a "status quo" analysis, as it represents the approach most commonly adopted in this literature.

The "status quo" results are presented in Figure 9. Panel A presents the results for math skills, and panel B presents the results for verbal skills. The results for domestic chore activities are presented in the left-hand side figures, while results for economic activities are on the right-hand side. A key result is that there is no evidence that domestic chores have a detrimental effect on either math or verbal scores. Only 1 of the 16 coefficients on chores in the math equation is significant and negative. The remaining coefficients are generally small in magnitude, and half of them are the "wrong" sign, with three of these even marginally significant (at the 10% level). In the verbal skills equation, the *only* significant coefficient on chores is the "wrong" sign, as are half the point estimates.

Recall that we found in Figure 5 that domestic chores had a significant negative effect on math scores if time was substituted from school or study to chores, but we found no significant effect if time was substituted from leisure to chores. And in Figure 6 we found that chores also had a significant negative effect on verbal scores if time was substituted from study. Thus, for chores, our analysis reveals a fundamentally different insight from the status quo analysis. While the status quo analysis produces a generic result that domestic chores are not detrimental for child cognitive development, we find that this is only true to the extent that chores substitute for leisure time. In contrast, our analysis reveals decidedly negative effects of chores to the extent that they substitute for school or study time.

The "status quo" results for economic activities are presented on the right-hand side of Figure 9. These results are the mirror image of the chores results. That is, the status quo analysis implies that economic activities are unambiguously detrimental for child math and verbal cognitive development. Yet we found this is not true if economic activities only substitutes for leisure time. Our analysis reveals a decidedly negative effect of economic activities on child development only if work time substitutes for school or study time.

Another group of "status quo" papers control for the actual hours that children work, but, like those cited earlier, do not control for time spent in other activities (see Akabayashi and Psacharopoulos 1999; Bezerra et al. 2009; Buonomo Zabaleta 2011; Heady 2003; Ray and Lancaster 2003). We replicate this type of analysis in Figure 10. Again, the results imply that hours spent on economic activities have an unambiguous negative effect on math and verbal skills. Conversely, there is no evidence that hours spent on domestic chores have a detrimental effect on verbal skills. The one contrast with Figure 9 is that in Figure 10 there is some evidence that hours spent on domestic chores is detrimental for math skills (i.e., 6 of 16 estimated effects are significantly negative), but this is still much weaker than the evidence for negative effects of economic activities (i.e., 14 of 16 effects significant).

Overall, therefore, the "status quo" analysis yields the conclusion that while economic activities are harmful for child development, time spent on domestic chores is either not harmful, or much less so. In contrast, we find that *both* time spent on domestic chores and economic activities are no less productive for child development than leisure time. But *both* time spent on domestic chores and economic activities are clearly less productive than time spent in school and (perhaps to an even greater degree) time spent studying.

This fundamental difference in the main take-away from the "status quo" approach relative to our approach can be explained by a combination of differences in allocation of non-work time among children engaged in economic activities relative to those doing domestic chores, as well as differences in the strength (if not the direction) of the effects of the other activities relative to work versus chores. Figures 1 and 2 help to explain the differences between the "status quo" and full time allocation vector results:

These figures show that, in most cases, time spent on economic activities is associated with greater reductions in school time than is time spent on domestic chores. This is particularly true at high levels of work or chore time (i.e., 6+ hours per day). Similarly, in most cases, a given amount of time spent on economic activities tends to be associated with a higher level of leisure than if the same amount of time is devoted to domestic chores. As a result of these patterns, a "status quo" analysis that fails to account for the whole time budget tends to find that work is much more detrimental for child development than chores, simply because work tends to be is associated with larger reductions in school time.

V.E. *Lagged Test Scores and Time Inputs*

Two key features of our extended value added model is that (i) we control for lagged test scores and lagged time inputs, and (ii) we allow the effects of these lagged inputs (as well as the coefficients on current inputs and background variables) to differ by both age and country. Table 5 reports the coefficients on lagged math scores in the math ability equations. Notice that the lagged math score, which the value added model uses to control for unobserved ability, is highly significant in all instances. At ages 12, 15 and 19, the lagged score coefficients range from 0.36 to 0.74, with a median point estimate of 0.46.

However, the coefficients on the age 5 score are much smaller in magnitude. As we noted in Section II.B, age 5 math skills were assessed using the CDA, an instrument designed for very young children, while math skills at later ages were assessed using paper and pencil tests. The changing nature of the test explains the smaller coefficient on lagged score at age 8.

Table 5: Coefficients on lagged test scores – MATH

	Ethiopia		India		Peru		Vietnam	
	coef	se	coef	Se	coef	se	coef	Se
Age 8: Lagged math score age 5	0.046**	0.017	0.154***	0.017	0.116***	0.014	0.117***	0.020
Age 12: Lagged math score age 8	0.558***	0.054	0.478***	0.028	0.532***	0.031	0.357***	0.041
Age 15: Lagged math score age 12	0.369***	0.040	0.425***	0.032	0.389***	0.031	0.378***	0.036
Age 19: Lagged math score age 15	0.550***	0.041	0.662***	0.047	0.736***	0.039	0.443***	0.039

Table 6 reports the coefficients on lagged verbal (PPVT test) scores in the verbal ability equations. Note that the lagged verbal score is highly significant in all instances. The lagged score coefficients range from 0.20 to 0.60, with a median point estimate of 0.31. Thus, PPVT is slightly less persistent over time than the math score.

Table 6: Coefficients on lagged test score – PPVT

	Ethiopia		India		Peru		Vietnam	
	coef	se	coef	Se	coef	se	coef	se
Age 8: lagged PPVT score age 5	0.204***	0.055	0.248***	0.044	0.386***	0.037	0.228***	0.035
Age 12: lagged PPVT score, age 8	0.326***	0.059	0.332***	0.031	0.486***	0.033	0.284***	0.051
Age 15: lagged PPVT score, age 12	0.218***	0.059	0.556***	0.044	0.597***	0.049	0.283***	0.046

The lagged time input coefficients are too numerous for us to report in detail. But we can highlight some key results. First, the lagged 5 time inputs are jointly significant (based on the F-test) in 12 of the 16 math equations (8 times at the 1% level, 4 times at the 5% level). Thus, including lagged inputs (as suggested by Todd and Wolpin, 2007) clearly improves the fit of the math ability equations. The lagged time inputs are somewhat less important in the PPVT equations. Specifically, they are jointly significant in only 4 of the 12 verbal ability equations. This is consistent with our finding that lagged test scores are also less important in the verbal equations than in the math equations, in the sense that lagged test scores are also a proxy for lagged inputs.

V.F. *Instrumenting for the Lagged Test Score and Time Use Measures*

As we discussed in Section IV, if test scores measure ability with error then the coefficient on the lagged score will tend to be biased downward, and the coefficients on time inputs are biased in an in an ambiguous direction. To address this issue, we re-estimated our main models using the lagged PPVT score to instrument for the lagged math score, and vice versa. We find that these are strong instruments, and the lagged test score coefficients increase as expected when we instrument (typically by 50% to 100% of the values reported in Tables 5 and 6). However, in the interest of space, in Figure 11 we report only the estimates of the child work coefficients with school time as the omitted category.

Figure 11 Panel A presents results for Math scores. These can be compared to our main results in Figure 5 Panel B. The results that instrument for the lagged math score using the lagged PPVT score are almost identical to the OLS estimates of the VA model. The point estimates of effects of domestic work and chores are nearly identical, and there is only a slight drop in precision of the estimates. For instance, in Figure 5 we found that 12 of the 16 estimates of the effect of economic activities relative to school were significantly negative at the 5% level, whereas in Figure 11 it is 11 out of 16. Similarly, in Figure 5 we found that 7 of the 16 estimates of the effect of domestic chores relative to school were significantly negative at the 5% level, and in Figure 11 we find 6 out of 16.[36]

Figure 11 panel B presents results for Verbal scores that can be compared to our main results in Figure 6 Panel B. The IV results again provide strong evidence that shifting children's time from school to economic activities is detrimental for development of verbal ability. Just as in Figure 5, we again find that 6 of the 12 estimates of the effect of economic activities relative to school are significantly negative. Also similar to Figure 5, the results are much weaker for domestic chores. Although 10 of the 12 point estimates are negative, only 1 is significant at the 5% level. In summary, it appears that instrumenting for the lagged test score makes little difference to our main results.

Finally, we consider the issue of possible measurement error in the parent reports of child time use. At ages 12 and 15, both parents and children were asked about the children's time use. Thus, in this subset of cases, we can use the child's reports to instrument for the parent's reports. However, as is well-known, IV estimators with several endogenous variables tend to have poor small sample properties. To help alleviate this problem, we grouped school and study time into a single "education time" category.

---

[36] Note that these are not the same 7 coefficients, as a couple of the coefficients shift between being significant at the 5% vs. 10% significance levels – in both directions.

Figure 12 reports IV and OLS results using "education time" as the omitted category. The OLS point estimates and confidence intervals are in black, while the IV results are in blue.[37] As is clear from the figure, while the IV estimates are slightly less precise, they are very similar to the OLS estimates. Consistent with our earlier findings, the IV results provide strong evidence that both economic activities and domestic chores have negative effects on both math and verbal scores (if they crowd out education time). The IV results also provide strong evidence that economic activities have negative effects on math scores, but, as before, the evidence for verbal scores is much weaker. Thus, instrumenting for child time use has little effect on our main findings.

VI. Conclusion

We have studied the trade-offs between the time children spend working (both on domestic chores and economic activities) and *specific* alternative times uses (e.g., leisure, school or study). This fills a critical gap in the existing child labour literature, which has focussed on the trade-off between work and a counterfactual bundle of alternative activities. Our current understanding of what would happen if children work less is, therefore, limited to the specific situation in which they reallocate the freed-up time to other activities in the same proportion as in the current bundle. But what we must know to design policies to promote child development and human capital accumulation is what reallocation would be most beneficial. Is any time use better than working, making reduction in the amount of time that children work a paramount aim? Or are some reallocations more productive than others, implying that policies should incentivise these specific reallocations? While the evidence available to date does not provide answers to these questions, they are highly pertinent to the huge ongoing policy effort to reduce the prevalence of child labour around the world.

We address this gap by estimating a child cognitive ability production function that accounts for the complete time budget of children – including time spent in economic activities, doing chores, attending school and studying at home. This allows us to identify more and less productive counterfactual time-use activities to child work. Our analysis is made possible by a unique multi-country longitudinal cohort study, Young Lives, which collects a wealth of data including information on the complete set of activities children engage in during a 24 hour period and state-of-the-art measures of cognitive skills.

[37] We do not report results for Ethiopia at age 12. Due to poor quality of the child time diary data, the child reports were very weak instruments for the parents' time reports.

The panel dimension of the data allows us to estimate value added production functions in which omitted inputs and unobserved ability are captured by both a lagged achievement measure and rich family background measures. We relax the assumptions of the standard value added model (i.e., input effects are age invariant and depreciate at a constant rate) by including controls for lagged inputs and letting effects of inputs differ by age.

Several key findings emerge. First, we find that across the four different countries (India, Vietnam, Peru and Ethiopia) and a wide age range (8, 12, 15 and 19) domestic chores and economic activities are not detrimental to cognitive development of children when compared to the activities they typically undertake during leisure time. However, we find clear evidence that both domestic chores and economic activities are detrimental to the development of cognitive skills if they crowd out school time. The detrimental effect of work time is even greater if it crowds out time spent studying at home.

Our approach yields substantively different findings from the standard approach of estimating the effect of child work without controlling for time spent in other activities. Had we done that, we would have concluded, as have many prior studies, that while time spent on economic activities is harmful for child development, time spent on chores is much less so. Such a finding would be consistent with the current policy focus of reducing children's engagement in economic activities *per se*.

Controlling for the full time-budget, however, what we find that both domestic chores and economic activities are detrimental to child development if they crowd out school and study time, but benign if they only crowd out leisure time. Thus, to be beneficial for child development and human capital accumulation, policy must focus on enabling children to spend more time at school and at study.

The difference in the conclusions drawn from the standard approach vs. our approach (which analyses the complete time budget) stems from differences in the typical counter-factual bundle of activities of children engaged in domestic chores relative to economic activities. On average, time spent on economic activities is associated with greater reductions in school time than time spent on domestic chores, especially at higher amounts of time.

In sum, the answer to the question "Is child work harmful for child cognitive development?" is "It depends." The key question is the extent to which child work conflicts with school attendance and studying at home. Policies that reduce child labour without also increasing schooling are unlikely to enhance human capital accumulation.

References

Akabayashi, Hideo and George Psacharopoulos. 1999. "The Trade-off between Child Labour and Human Capital Formation: A Tanzanian Case Study." *The Journal of Development Studies* 35(5):120–140.

Amin, Shahina, Shakil Quayes, and Janet M. Rives. 2006. "Market Work and Household Work as Deterrents to Schooling in Bangladesh." *World Development* 34(7):1271–86.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5(4):1–27.

Assad, R., D. Levison, and H. Dang. 2010. "How Much Work Is Too Much? Effect of Child Work on Schooling - the Case of Egypt"." Pp. 53–97 in *Child Labor and the Transition between School and Work*, edited by R. Akee, E. Edmonds, and K. Tatsiramos. Bingley:Emerald.

Attanasio, Orazio et al. 2010. "Children's Schooling and Work in the Presence of a Conditional Cash Transfer Program in Rural Colombia." *Economic Development and Cultural Change* 58(2):181–210.

Attanasio, Orazio, Costas Meghir, Emily Nix, and Francesca Salvati. 2017. "Human Capital Growth and Poverty: Evidence from Ethiopia and Peru." *Review of Economic Dynamics* 25:234–59.

Banerjee, Abhijit V., Swati Bhattacharjee, R. Chattopadhyay, and Alejandro J. Ganimian. 2017. *The Untapped Math Skills of Working Children in India: Evidence, Possible Explanations, and Implications.*

Basu, Kaushik. 1999. "Child Labor: Cause, Consequence, and Cure, with Remarks on International Labor Standards." *Journal of Economic Literature* 37(3):1083–1119.

Beegle, Kathleen, Rajeev Dehejia, and Roberta Gatti. 2009. "Why Should We Care About Child Labor?: The Education, Labor Market, and Health Consequences of Child Labor." *Journal of Human Resources* 44(4):871–89.

Bezerra, M. E. G., A. L. Kassouf, and M. A. Kuenning. 2009. *The Impact of Child Labor and School Quality on Academic Achievement in Brazil*. 4062.

Del Boca, Daniela, Chiara Monfardini, and Cheti Nicoletti. 2016. "Parental and Child Time Investments and the Cognitive Development of Adolescents." *Journal of Labor Economics* 35(2):565–608.

Buonomo Zabaleta, Mariela. 2011. "The Impact of Child Labor on Schooling Outcomes in Nicaragua." *Economics of Education Review* 30(6):1527–39.

Campbell, J. M., S. K. Bell, and L. K. Keith. 2001. "Concurrent Validity of the Peabody Picture Vocabulary Test-Third Edition As an Intelligence and Achievement Screener for Low SES African American Children." *Assessment* 8(1):85–94.

Campbell, Jonathan. 1998. "Book Review: Peabody Picture Vocabulary Test, Third Edition." *Journal of Psychoeducational Assessment* 16(4):334–38.

Cigno, Alessandro and Furio C. Rosati. 2005. *The Economics of Child Labour*. Oxford University Press.

Das, Jishnu and Tristan Zajonc. 2010. “India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement.” *Journal of Development Economics* 92(2):175–87.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. “School Choice, School Quality, and Postsecondary Attainment.” *The American Economic Review* 104(3):991–1013.

Dumas, Christelle. 2012. “Does Work Impede Child Learning? The Case of Senegal.” *Economic Development and Cultural Change* 60(4):773–93.

Dunn, L. M., L. M. Dunn, and American Guidance Service. 1997. *A Examiner’s Manual for the PPVT-III, Peabody Picture Vocabulary Test, Third Edition*. AGS.

Dunn, L., E. Padilla, D. Lugo, and L. Dunn. 1986. *Manual Del Examinador Para El Test de Vocabularo En Imagenes Peabody: Adaptacion Hispanoamericana (Peabody Picture Vocabulary Test: Hispanic-American Adaptation)*. Minnesota: AGS.

Edmonds, Eric V. 2007. *Chapter 57 Child Labor*. Vol. 4. Elsevier B.V.

Edmonds, Eric V. 2009. *Defining Child Labour : A Review of the Definitions of Child Labour in Policy Research*.

Emerson, Patrick M., Vladimir Ponczek, and André Portela Souza. 2017. “Child Labor and Learning.” *Economic Development and Cultural Change* 65(2):265–96.

Emerson, Patrick M. and André Portela Souza. 2011. “Is Child Labor Harmful? The Impact of Working Earlier in Life on Adult Earnings.” *Economic Development and Cultural Change* 59(2):345–85.

Glewwe, P. and K. Muralidharan. 2016. “Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications.” Pp. 653–743 in Vol. 5, edited by E. A. Hanushek, S. Machin, and L. B. T.-H. of the E. of E. Woessmann. Elsevier.

Glewwe, Paul. 2002. “Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes.” *Journal of Economic Literature* 40(2):436–82.

Glewwe, Paul and Michael Kremer. 2006. “Chapter 16 Schools, Teachers, and Education Outcomes in Developing Countries.” Pp. 945–1017 in Vol. 2, edited by E. Hanushek and F. B. T.-H. of the E. of E. Welch. Elsevier.

Gray, Shelley, Elena Plante, Rebecca Vance, and Mary Henrichsen. 1999. “The Diagnostic Accuracy of Four Vocabulary Tests Administered to Preschool-Age Children.” *Language, Speech, and Hearing Services in Schools* 30(2):196–206.

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. 2014. “Can Value-Added Measures of Teacher Performance Be Trusted?” *Education Finance and Policy* 10(1):117–56.

Gunnarsson, Victoria, Peter F. Orazem, and Mario A. Sánchez. 2006. "Child Labor and School Achievement in Latin America." *The World Bank Economic Review* 20(1):31–54.

Hanushek, Eric A. and Dennis D. Kimko. 2000. "Schooling, Labor-Force Quality, and the Growth of Nations." *The American Economic Review* 90(5):1184–1208.

Hanushek, Eric A. and Lei Zhang. 2009. "Quality Consistent Estimates of International Schooling and Skill Gradients." *Journal of Human Capital* 3(2):107–43.

Heady, Christopher. 2003. "The Effect of Child Labor on Learning Achievement." *World Development* 31(2):385–98.

ILO. 2017. *Global Estimates of Child Labour: Results and Trends, 2012-2016*.

International Labour Organization. 2017. *Ending Child Labour by 2025: A Review of Policies and Programmes*.

J Heckman, J., J. Stixrud, and S. Urzua. 2006. *The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Socialbehavior*. Vol. 24.

Jacoby, Hanan G. and Emmanuel Skoufias. 1997. "Risk, Financial Markets, and Human Capital in a Developing Country." *The Review of Economic Studies* 64(3):311–35.

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*, edited by W. J. van der Linden and R. K. Hambleton. Springer, New York, N.Y.

Lindqvist, Erik and Roine Vestman. 2011. "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment." *American Economic Journal: Applied Economics* 3(1):101–28.

Montie, Jeanne E., Zongping Xiang, and Lawrence J. Schweinhart. 2006. "Preschool Experience in 10 Countries: Cognitive and Language Performance at Age 7." *Early Childhood Research Quarterly* 21(3):313–31.

Muralidharan, K. and S. Venkatesh. 2013. *Contract Teachers: Experimental Evidence from India*. 19440.

Orazem, Peter and L. Victoria Gunnarsson. 2004. *Child Labour, School Attendance and Performance: A Review*. 04001.

Patrinos, Harry Anthony and George Psacharopoulos. 1997. "Family Size, Schooling and Child Labor in Peru--An Empirical Analysis." *Journal of Population Economics* 10(4):387–405.

Psacharopoulos, George. 1997. "Child Labor versus Educational Attainment Some Evidence from Latin America." *Journal of Population Economics* 10(4):377–86.

Rahman, Mohammad Mafizur, Rasheda Khanam, and Nur Uddin Absar. 1999. "Child Labor in Bangladesh: A Critical Appraisal of Harkin's Bill and the MOU-Type Schooling Program." *Journal of Economic Issues* 33(4):985–1003.

Rao, Nirmala et al. 2014. “Early Childhood Development and Cognitive Development in Developing Countries.” *Education Rigorous Literature Review* (September).

Ravallion, Martin and Quentin Wodon. 2000. “Does Child Labour Displace Schooling? Evidence on Behavioural Responses to an Enrollment Subsidy.” *The Economic Journal* 110(462):C158–75.

Ray, Ranjan and Geoffrey Lancaster. 2003. “Does Child Labour Affect School Attendance and School Performance?Multi Country Evidence on SIMPOC Data.” *Econometric Society 2004 Australasian Meetings* (July):68.

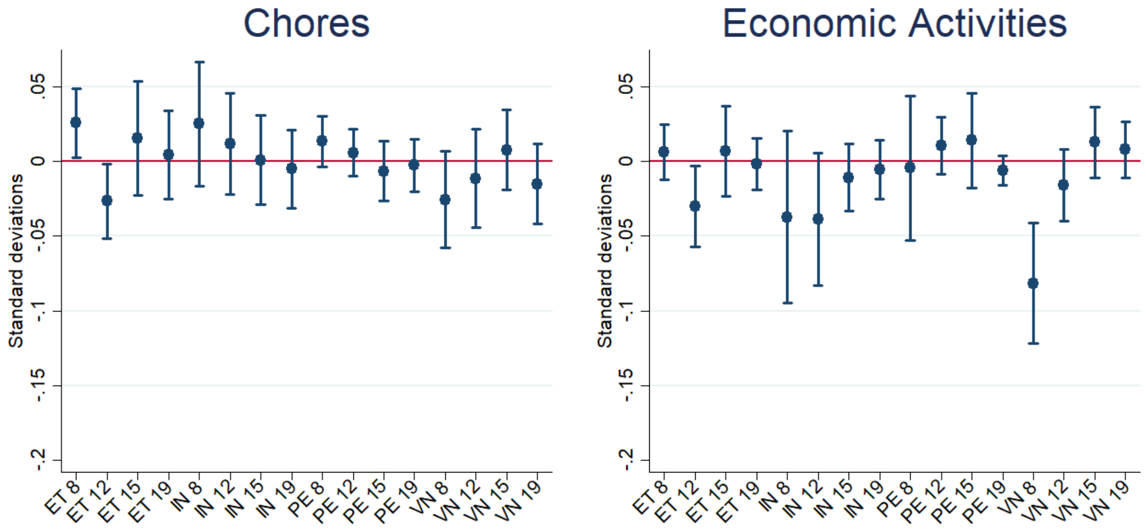
Sedlacek, G., S. Duryea, N. Ilahi, and M. Sasaki. 2009. “Child Labor, Schooling and Poverty in Latin America.” in *Child Labor and Education in Latin America*, edited by P. F. Orazem, G. Sedlacek, and Z. Tzannatos. Palgrave Macmillan, New York.

Singh, Abhijeet. 2014. *Emergence and Evolution of Learning Gaps across Countries*. 124.

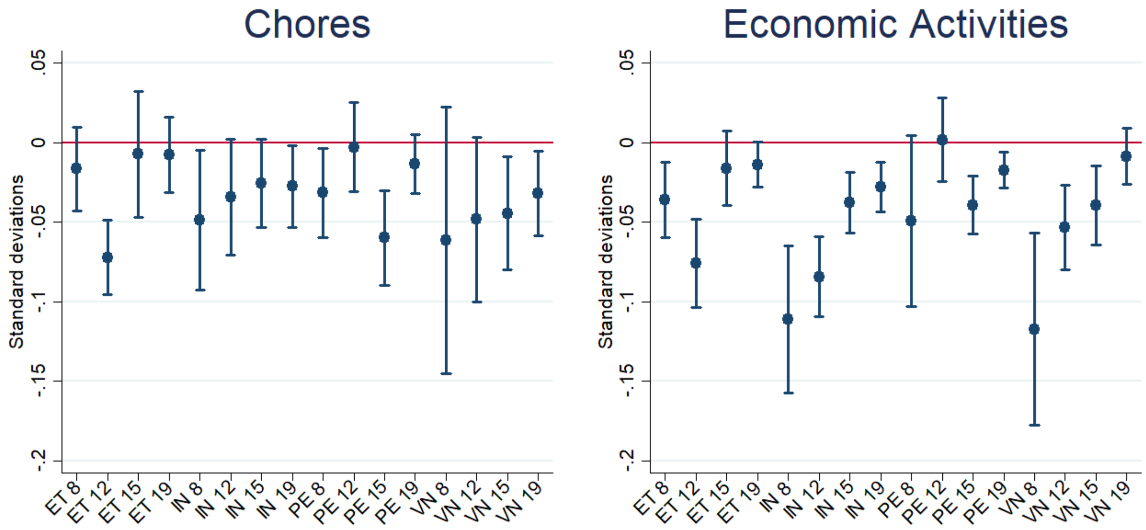
Todd, P. and K. Wolpin. 2007. “The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps.” *Journal of Human Capital* 1(1):91–136.

Figure 5: Effect of Child Work on Math Scores

Panel A – Leisure Omitted



Panel B – School Omitted



Panel C – Study Omitted

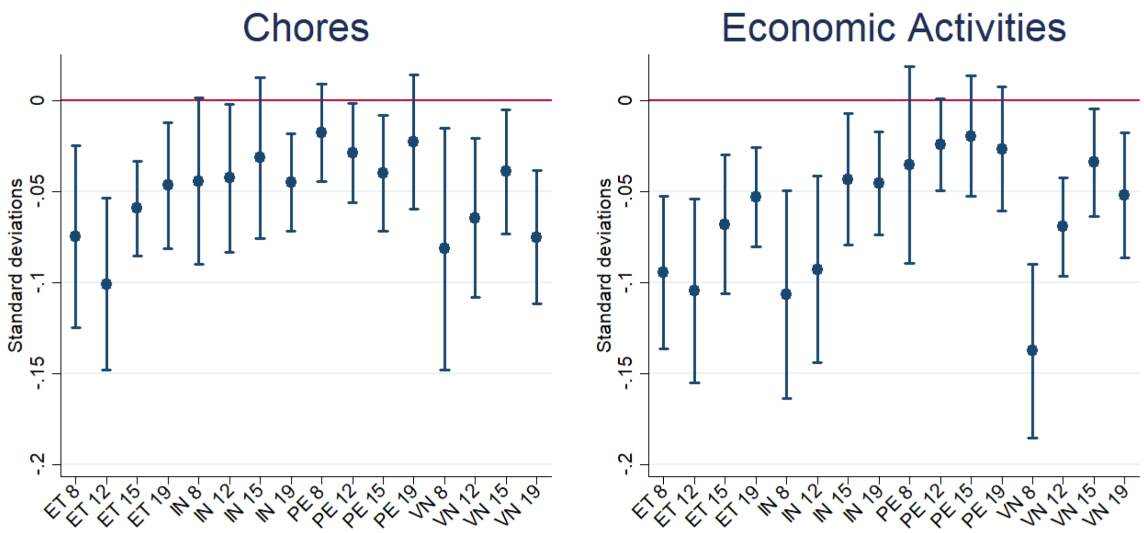
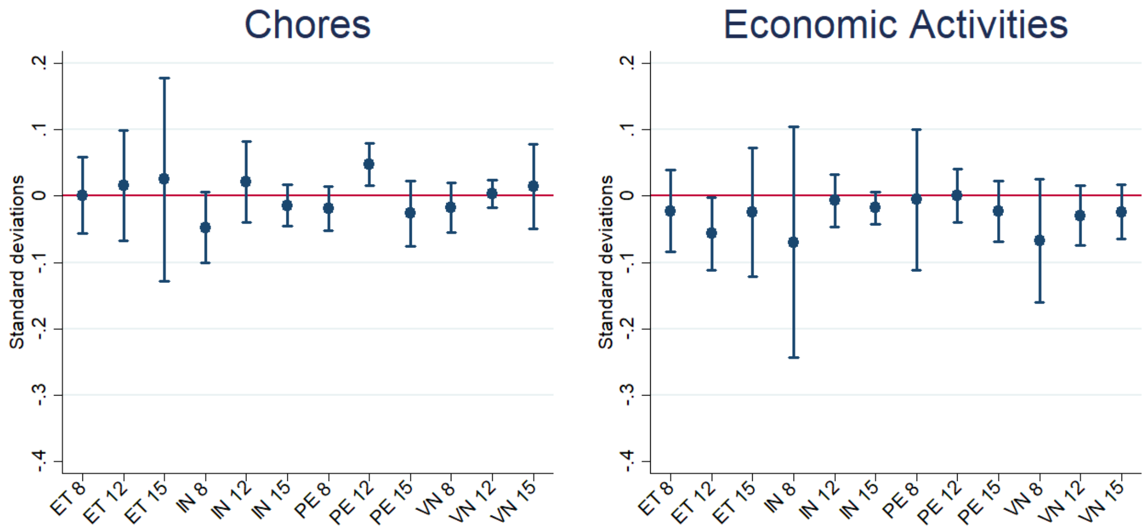
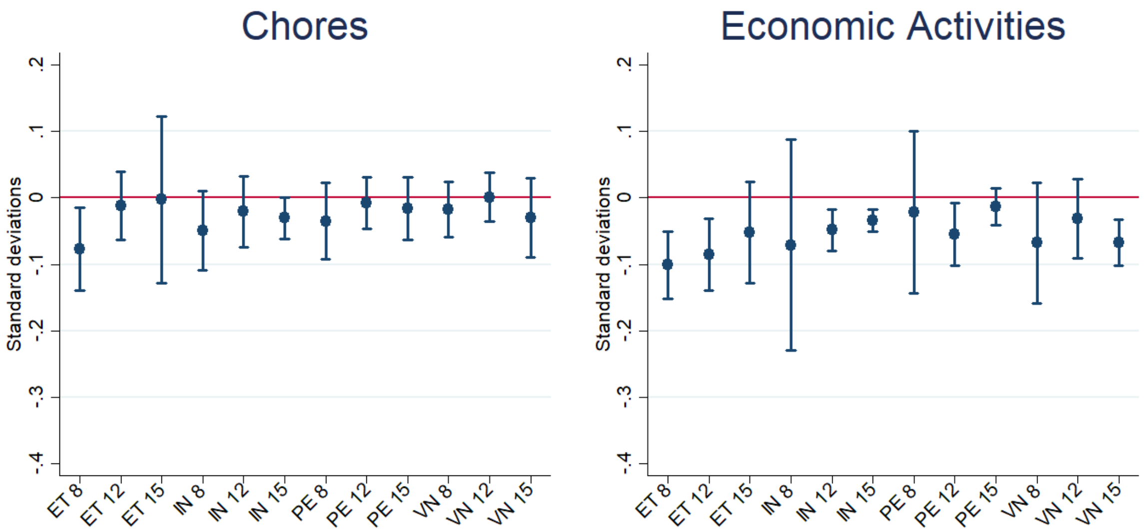


Figure 6: Effect of Child Work on Verbal Scores

Panel A – Leisure Omitted



Panel B – School Omitted



Panel C – Study Omitted

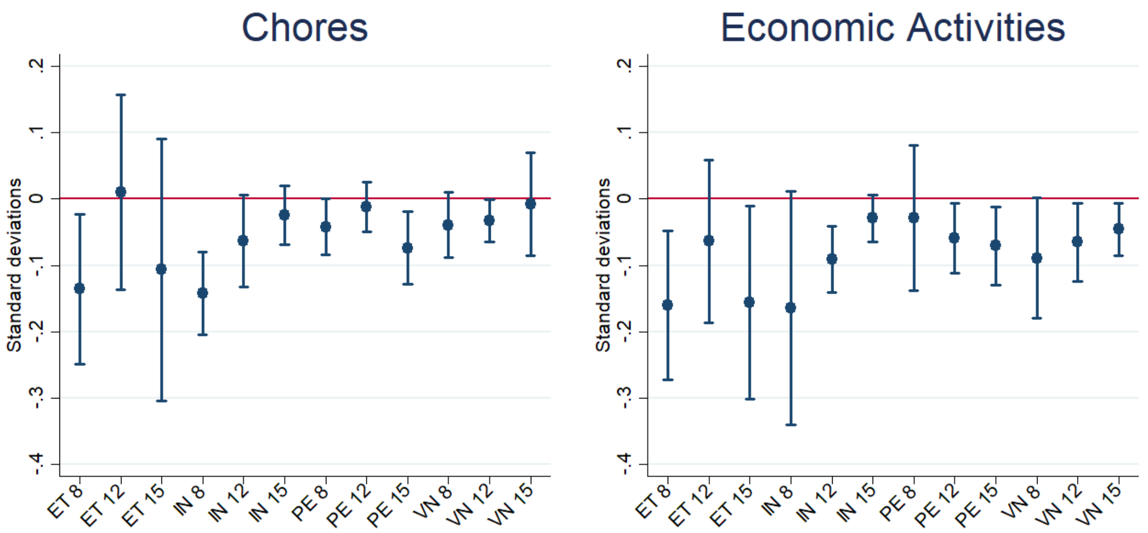


Figure 7: Effect of Schooling on Math and Verbal scores (relative to Study)

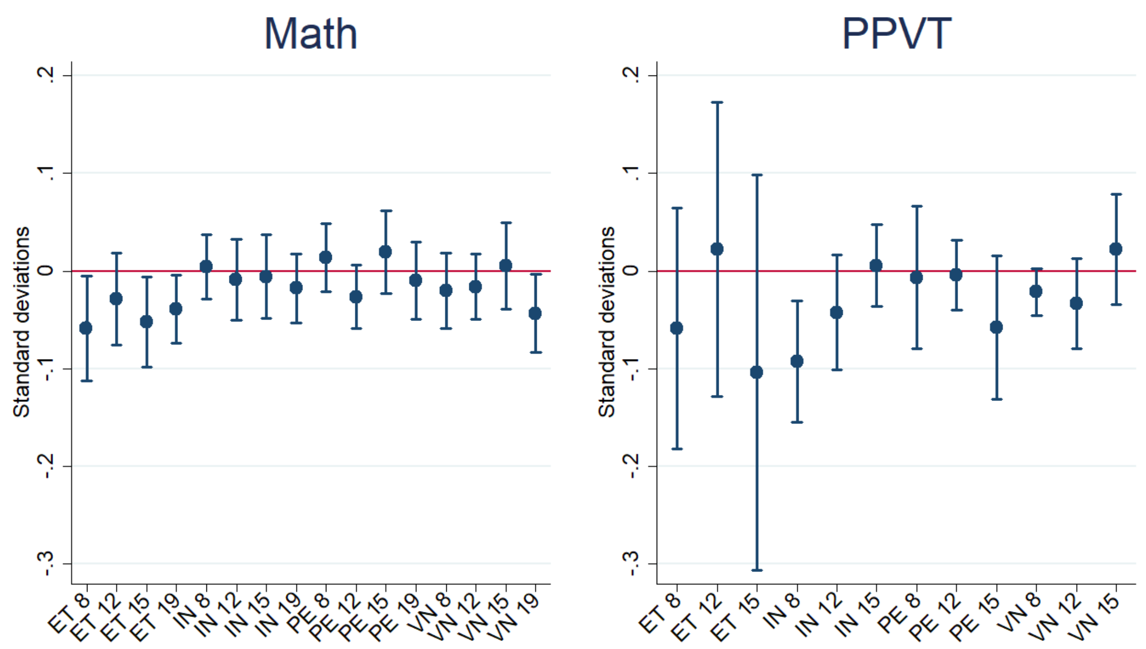


Figure 8: Effect of Economic Activities on Math and Verbal scores (relative to Chores)

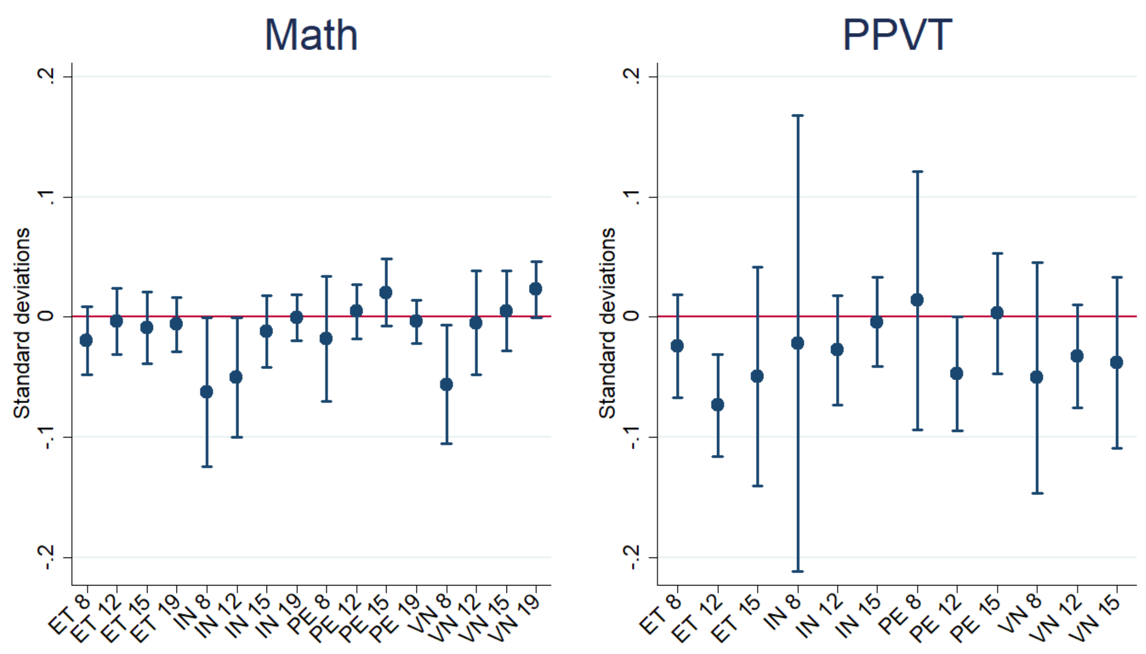
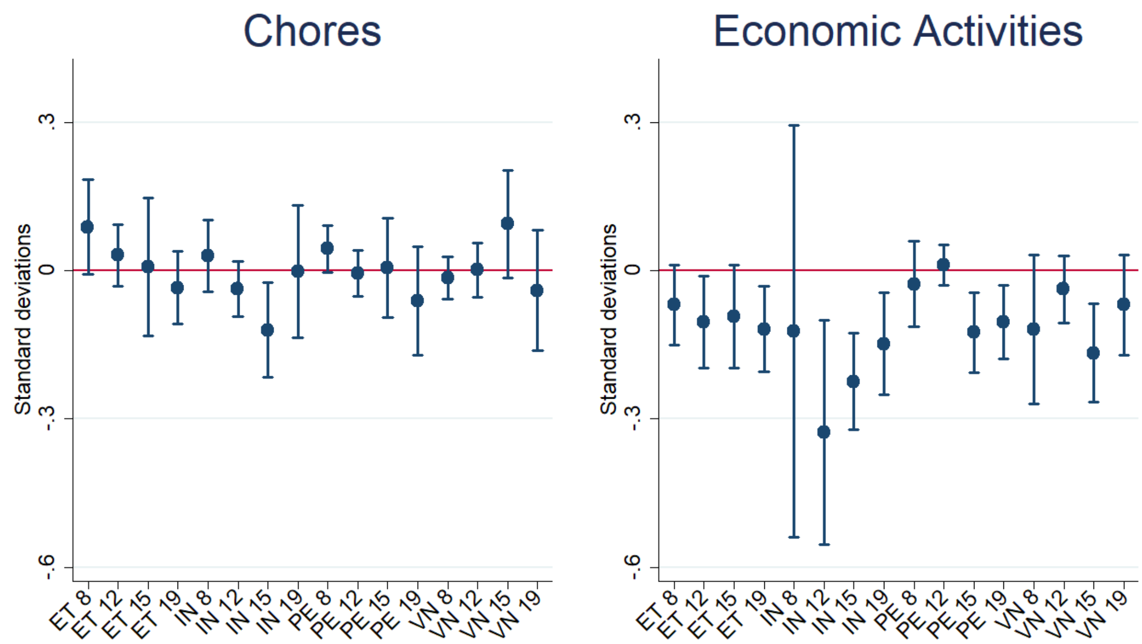
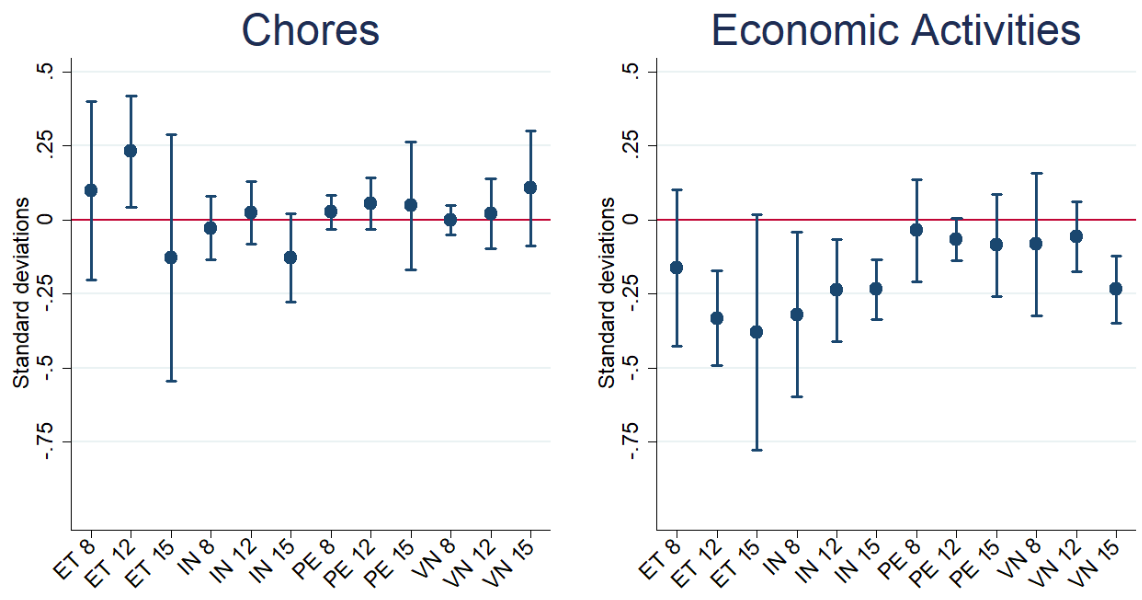


Figure 9: "Status Quo" Analysis of Impact of Child Work on Math and Verbal scores

Panel A – Math Skills



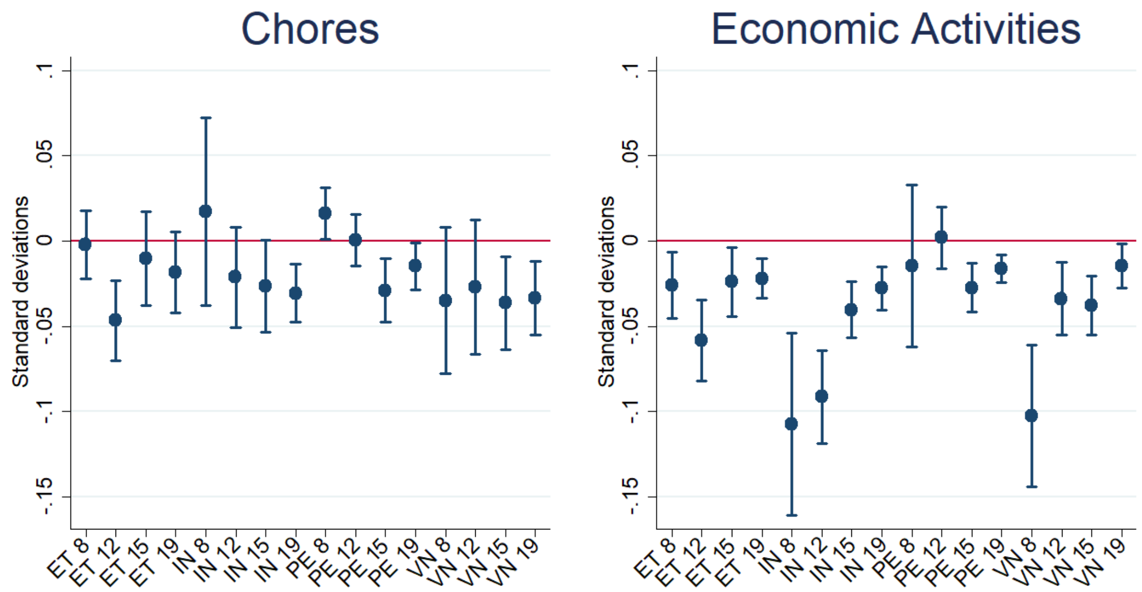
Panel B – Verbal Skills



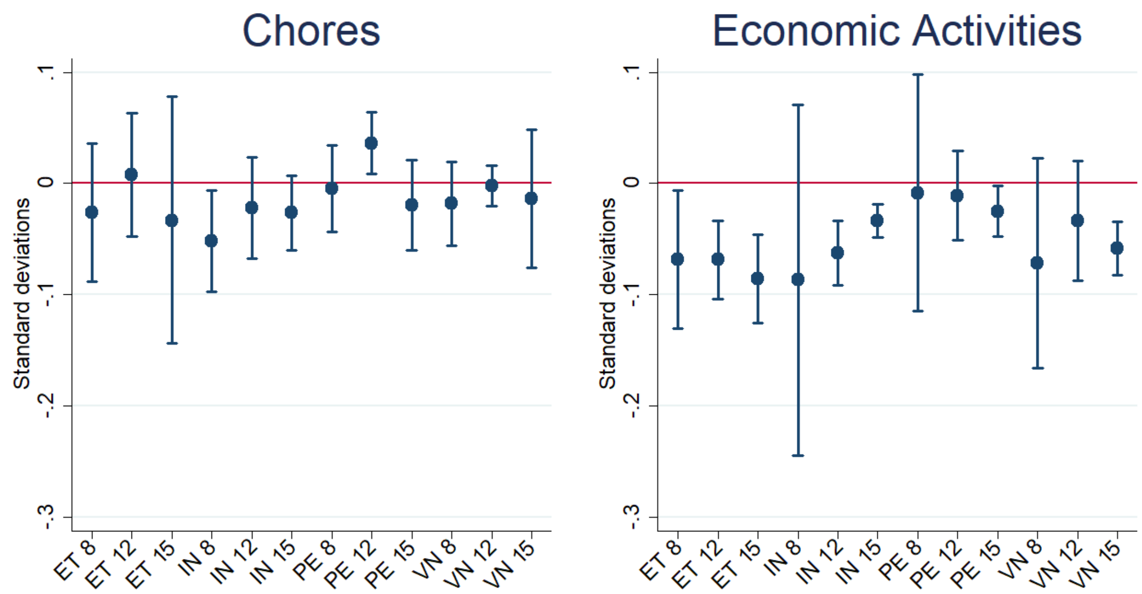
Note: Coefficients are from regressions that control only for indicators of whether the child engages in any economic activities/household chores. This "status quo" approach contrasts with our main models that control for time allocated to six possible time-use categories that make up a complete 24 hours.

Figure 10: “Status Quo” Analysis of Impact of Work Hours on Math and Verbal scores

Panel A – Math Skills



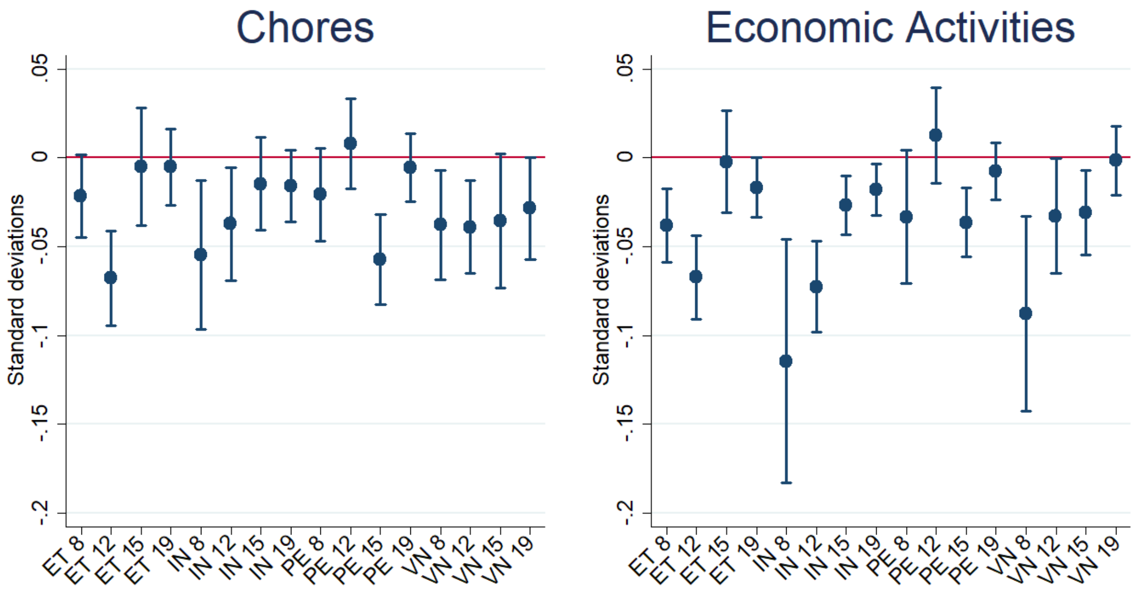
Panel B – Verbal Skills



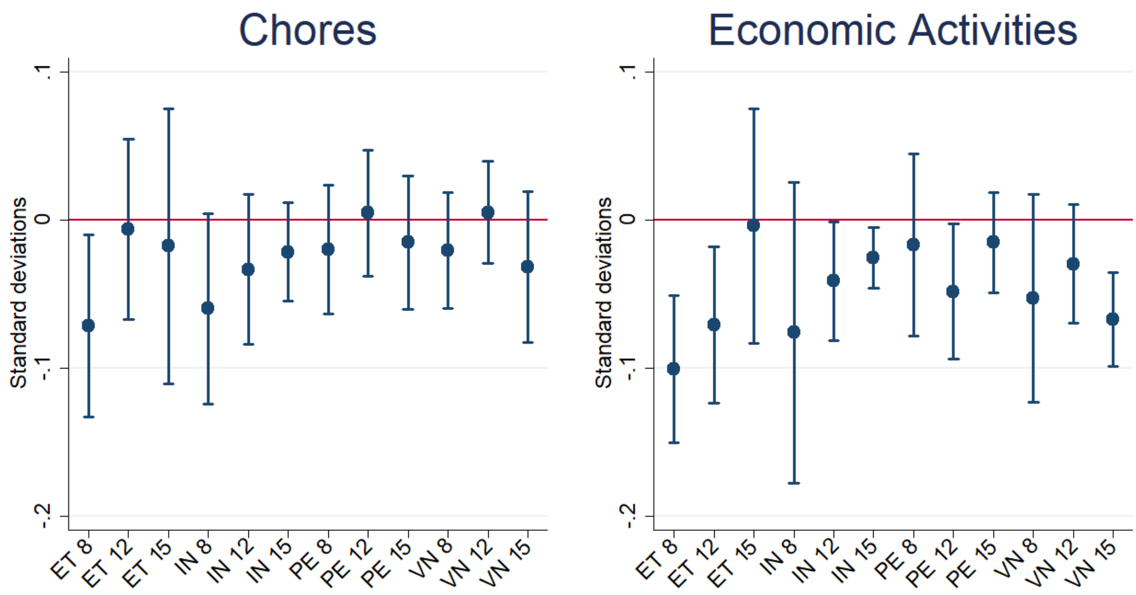
Note: Coefficients are from regressions that control only for hours of work in economic activities and in household chores. This “status quo” approach contrasts with our main models that control for time allocated to six possible time-use categories that make up a complete 24 hours.

Figure 11: Effect of Child Work on Math and Verbal scores (IV for lagged test score)

Panel A – Math Skills



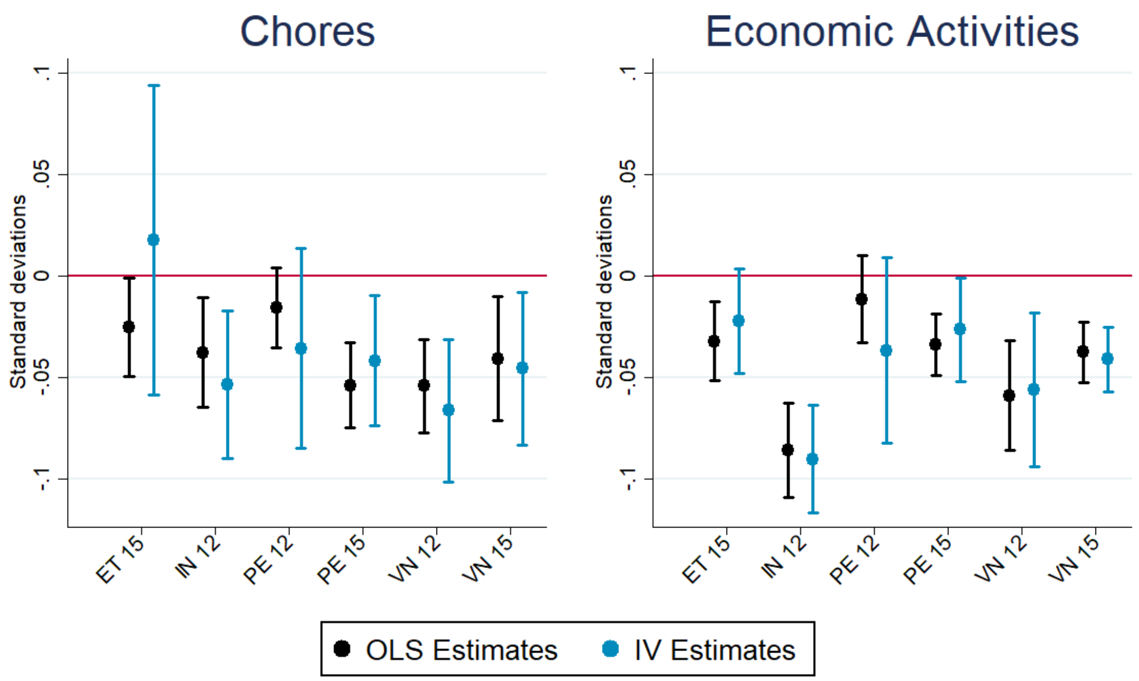
Panel B – Verbal Skills



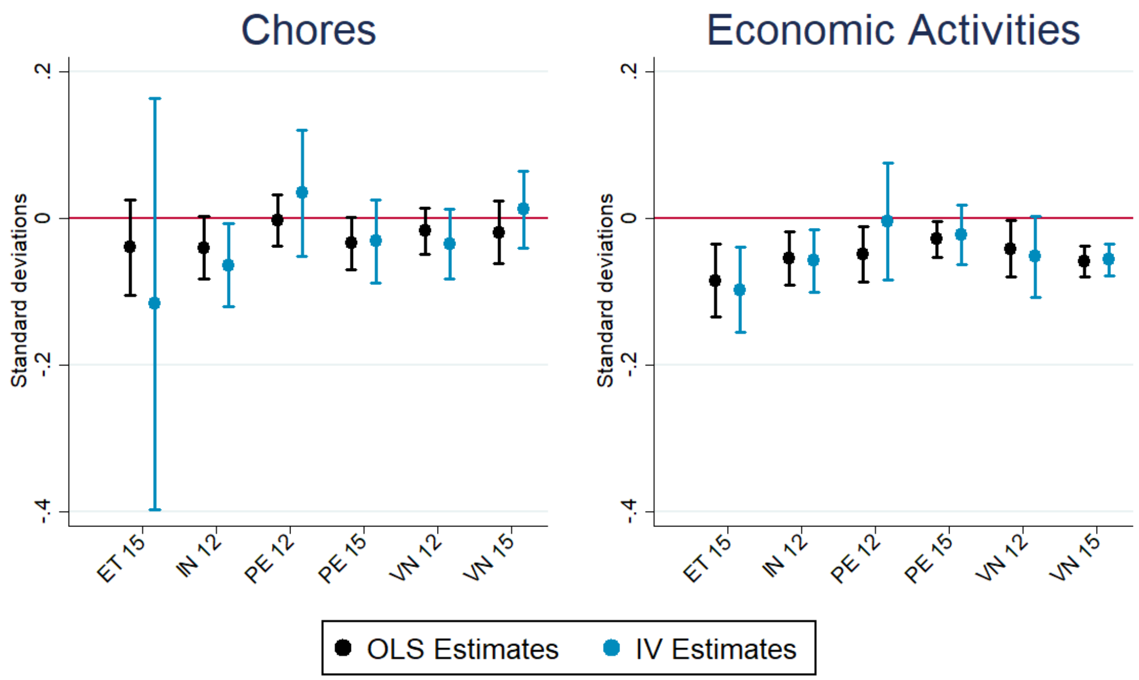
Note: Schooling time is the omitted time use category in this figure. We use the lagged verbal score to instrument for the lagged math score, and vice versa.

Figure 12: Effect of Work on Math and Verbal scores (OLS vs IV for time allocation)

Panel A – Math Skills



Panel B – Verbal Skills



Note: Education time (defined as school plus study time) is the omitted time use category in this figure (The four included time use categories are domestic chores, economic activities, leisure and sleep). For the IV estimates we use the child's self-reported time-use to instrument for the parent reports of child time use. The child reports are only available at ages 12 and 15.

Appendix Table 1: Summary statistics

	Ethiopia		India		Peru		Vietnam	
	YC (age 5)	OC (age 12)	YC (age 5)	OC (age 12)	YC (age 5)	OC (age 12)	YC (age 5)	OC (age 12)
Male	0.528	0.509	0.534	0.495	0.503	0.533	0.512	0.496
	(0.499)	(0.500)	(0.499)	(0.500)	(0.500)	(0.499)	(0.500)	(0.500)
Mother's Education:								
None	0.496	0.442	0.512	0.582	0.0829	0.101	0.116	0.0954
	(0.500)	(0.497)	(0.500)	(0.493)	(0.276)	(0.301)	(0.320)	(0.294)
primary	0.397	0.485	0.279	0.289	0.454	0.506	0.484	0.493
	(0.489)	(0.500)	(0.449)	(0.454)	(0.498)	(0.500)	(0.500)	(0.500)
secondary	0.0869	0.0616	0.177	0.103	0.279	0.247	0.327	0.363
	(0.282)	(0.241)	(0.382)	(0.304)	(0.448)	(0.432)	(0.469)	(0.481)
post secondary	0.0197	0.0117	0.0321	0.0259	0.184	0.146	0.0730	0.0488
	(0.139)	(0.108)	(0.176)	(0.159)	(0.388)	(0.353)	(0.260)	(0.216)
Father's Education:								
none	0.209	0.171	0.333	0.397	0.0115	0.0136	0.0761	0.0679
	(0.407)	(0.377)	(0.471)	(0.490)	(0.107)	(0.116)	(0.265)	(0.252)
primary	0.628	0.715	0.320	0.333	0.346	0.563	0.460	0.446
	(0.483)	(0.452)	(0.467)	(0.472)	(0.476)	(0.496)	(0.499)	(0.497)
secondary	0.114	0.0850	0.255	0.201	0.452	0.274	0.378	0.406
	(0.318)	(0.279)	(0.436)	(0.401)	(0.498)	(0.446)	(0.485)	(0.491)
post secondary	0.0487	0.0287	0.0925	0.0685	0.190	0.149	0.0855	0.0795
	(0.215)	(0.167)	(0.290)	(0.253)	(0.393)	(0.356)	(0.280)	(0.271)
Household size	6.038	6.532	5.521	5.182	5.490	5.560	4.673	4.900
	(2.058)	(2.041)	(2.232)	(1.826)	(2.072)	(1.970)	(1.514)	(1.391)
No. brothers in house	1.454	1.885	0.653	0.968	0.953	1.286	0.536	0.887
	(1.380)	(1.448)	(0.725)	(0.842)	(1.118)	(1.161)	(0.707)	(0.846)
No. sisters in house	1.384	1.861	0.810	0.890	0.951	1.127	0.616	0.881
	(1.333)	(1.411)	(0.875)	(0.901)	(1.153)	(1.067)	(0.842)	(0.972)
No. grandparents in house	0.238	0.181	0.959	0.507	0.617	0.346	0.680	0.278
	(0.548)	(0.467)	(0.963)	(0.752)	(0.891)	(0.674)	(0.907)	(0.564)
No. other adults in house	0.419	0.407	0.764	0.341	0.709	0.475	0.452	0.148
	(1.067)	(1.012)	(1.424)	(1.054)	(1.271)	(1.044)	(1.003)	(0.542)
No. other elderly in house	0.00547	0.0106	0.00105	0	0.0132	0.0121	0.00573	0.00212
	(0.0738)	(0.103)	(0.0324)	(0)	(0.127)	(0.109)	(0.0755)	(0.0460)
Both parents in house	0.759	0.651	0.945	0.863	0.791	0.709	0.919	0.917
	(0.428)	(0.477)	(0.228)	(0.344)	(0.407)	(0.454)	(0.273)	(0.276)
Mother's age	31.45	38.26	27.63	34.70	31.27	38.40	31.18	38.43
	(6.404)	(6.946)	(4.315)	(5.634)	(6.566)	(6.555)	(5.782)	(5.713)
Father's age	40.72	47.95	33.44	40.89	35.31	42.47	34.09	40.77
	(8.650)	(8.390)	(5.203)	(6.160)	(6.889)	(6.942)	(5.948)	(6.025)

Child's age (in months)	62.37	145.2	64.74	148.5	63.98	148.2	63.67	147.6
	(3.796)	(3.718)	(3.713)	(4.080)	(4.681)	(5.227)	(3.619)	(3.782)
Child lives in urban area	0.401	0.407	0.252	0.245	0.695	0.747	0.205	0.196
	(0.490)	(0.492)	(0.434)	(0.431)	(0.461)	(0.435)	(0.404)	(0.397)
Height for age z-score	-1.441	-1.374	-1.638	-1.641	-1.532	-1.531	-1.338	-1.472
	(1.121)	(1.264)	(1.111)	(1.672)	(1.131)	(1.157)	(1.109)	(1.088)
Wealth index	0.287	0.303	0.459	0.469	0.470	0.504	0.490	0.512
	(0.178)	(0.169)	(0.195)	(0.200)	(0.230)	(0.222)	(0.181)	(0.171)
Child Religion (1)	0.717	0.727	0.875	0.874	0.809	0.837	0.857	0.839
	(0.451)	(0.446)	(0.330)	(0.331)	(0.393)	(0.369)	(0.350)	(0.368)
Child Religion (2)	0.114	0.112	0.0699	0.0674	0.133	0.133	0.143	0.161
	(0.318)	(0.315)	(0.255)	(0.251)	(0.340)	(0.339)	(0.350)	(0.368)
Child Religion (3)	0.160	0.154	0.0547	0.0581	0.0577	0.0301	NA	NA
	(0.366)	(0.361)	(0.227)	(0.234)	(0.233)	(0.171)		
Child Religion (4)	0.00984	0.00744	NA	NA	NA	NA	NA	NA
	(0.0987)	(0.0860)						
Child Ethnicity (1)	0.288	0.287	0.181	0.206	0.916	0.926	0.857	0.874
	(0.453)	(0.453)	(0.385)	(0.405)	(0.277)	(0.262)	(0.350)	(0.332)
Child Ethnicity (2)	0.214	0.207	0.148	0.109	0.0561	0.0422	0.143	0.126
	(0.410)	(0.406)	(0.355)	(0.312)	(0.230)	(0.201)	(0.350)	(0.332)
Child Ethnicity (3)	0.228	0.228	0.469	0.469	0.0278	0.0316	NA	NA
	(0.420)	(0.420)	(0.499)	(0.499)	(0.164)	(0.175)		
Child Ethnicity (4)	0.270	0.277	0.139	0.152	NA	NA	NA	NA
	(0.444)	(0.448)	(0.346)	(0.360)				
Child Ethnicity (5)	NA	NA	0.0636	0.0633	NA	NA	NA	NA
			(0.244)	(0.244)				
N	1829	941	1903	964	1906	664	1919	943

Notes: Standard deviations in brackets; Sample includes all those for whom at least one of the main models can be estimated. Height-for-age z-scores calculated using WHO 2006 reference tables. Religion codes by country: Ethiopia 1=Christian Orthodox, 2=Other Christian, 3=Muslim, 4=Other; India 1=Hindu, 2=Muslim, 3=Other (includes Christian, Buddhist); Peru 1=Catholic, 2=Evangelist, 3=Other (biggest group = none); Vietnam 1=none, 2=Other (biggest groups include Buddhist, ancestor worship). Ethnicity codes by country: Ethiopia 1=Amhara, 2=Oromo, 3=Tigrian, 4=Other (biggest groups include Gurage, Hadiva, Sidama, Wolavta); India 1= Scheduled Caste, 2=Scheduled Tribe, 3=Backward Caste, 4=Other Hindu, 5=Other non-Hindu; Peru 1=Mestizo, 2=White, 3=Other; Vietnam 1=Majority (Kinh), 2=Minority (biggest groups include H'mong, Dao, Tay, Nung). Wealth index, constructed and publicly archived by the Young Lives team is a simple average of three separate indexes that range from 0 to 1: housing quality, consumer durables, and access to services. The housing quality index is a mean of (1) rooms per person (number of rooms divided by number of household members), set to take a maximum value of 1; (2) floor quality (a dummy variable which takes the value of 1 if the floor is made of finished material); and (3) roof quality (a dummy variable that takes the value of 1 if the roof is made of iron, concrete tiles, or slate. The consumer durables index is the proportion of durables a household owns from a list of seven (radio, motorbike/scooter, bicycle, TV, motorized vehicle or truck, landline telephone, modern bed or table). The services index is the proportion of key services that a household has access to: electricity, piped water, own pit latrine/flush toilet, and modern cooking fuel (gas, kerosene, or electricity).