IFS

Michael Keane
Sonya Krutikova
Timothy Neal

20/36

Working paper

The impact of child work on cognitive development: results from four low to middle income countries

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]

October 14, 2020

Abstract: We study the impact of child work on cognitive development in four Low- and Middle-Income Countries. We advance the literature by using cognitive test scores collected regardless of school attendance. We also address a key gap in the literature by controlling for children's complete time allocation budget. This allows us to estimate effects of different types of work, like chores and market/farm work, relative to specific alternative time-uses, like school or study or play/leisure. Our results show child work is more detrimental to child development to the extent that it crowds out school/study time rather than leisure. We also show the adverse effect of time spent on domestic chores is similar to time spent on market and farm work, provided they both crowd out school/study time. 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 labor; Child development; Education; Time use; Item response theory; Value added models

---

[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 labor harmful for child development? Existing evidence suggests the answer is "probably yes," but prior results are inconclusive. A key problem is that existing studies of child work do not control for how children spend their non-work time. Hence, we lack a good understanding of the trade-offs between child work and other uses of time – such as school, leisure and chores. Instead, prior studies reveal the trade-off between child work and some un-specified *bundle* of alternative activities. Such estimates are of limited use for directing policy, because they only tell us what would happen if time freed up via reduced child work is reallocated to other activities in the same proportions as in the current bundle (an effect that is context specific). Prior estimates are not informative about the effects of policies that aim to shift work time to a *specific* alternative, such as enhanced school 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 examine trade-offs between child work and the other alternative time-uses.[4] We find that whether child work is harmful depends crucially on what the child would do instead. Policies to reduce child work will only increase human capital if they incentivize families to reallocate the freed-up time to more productive activities, such as school/study.

Our analysis is made possible by Young Lives, a multi-country panel dataset that follows children in four low to middle income countries – Ethiopia, India (Andhra Pradesh and Telengana), Vietnam and Peru – from birth until age 22. The data contain information on time use over a "typical" 24 hour period, including time at school, studying at home, doing paid work, chores and farm work, caring for others, sleeping and playing. The survey also uses state-of-the-art methods to measure children's cognitive skills using scores on math and verbal tests. For the first time in the child labor literature, the Young Lives survey collected test scores that are *both* comparable over time and obtained regardless of school attendance.

We use Young Lives data to estimate child cognitive ability production functions that include a complete vector of time inputs, as well as rich controls for other inputs (i.e., parent education, wealth, family structure). Of course, we cannot measure all possible inputs to child development, or control perfectly for latent child ability. But we exploit the panel aspect of

---

[4] We define "child work" to include not only paid work outside the household, but also domestic chores and care responsibilities, farm tasks and work on the family farm/business. Child labour is typically defined as a narrower set of activities which excludes domestic chores (see International Labour Organization 2017). Hence we adopt the broader term "child work" to describe all these activities.

the data to estimate value added (VA) production functions where omitted inputs and latent ability are proxied by a lagged test score (as in Todd and Wolpin, 2007, Fiorini and Keane, 2014). Several recent studies find VA models are a reliable way to control for latent ability.[5]

A key finding is that child work (including both market work and household chores) has little or no adverse effect on cognitive development if it only substitutes for leisure. The implication is that policies that merely shift work time to leisure will do little to enhance child cognitive development. In contrast, we find school and study time are far more productive than child work. Thus, if one's goal is to enhance child development, it is important to design policies that not only reduce child work but also shift time to educational activities.

We also show how the typical or "status quo" approach in the literature of estimating effects of child labor without control for children's complete time-use budget can generate misleading conclusions. In particular, prior literature typically finds market and farm work are much more harmful for child development than household chores. We find this is largely because market and farm work tend to crowd out more school/study time than do chores. However, we find that chores that crowd out school time are almost as detrimental as market and farm work – a result that has been masked in prior analysis.

Our VA results are robust to using fixed effects rather than VA models. We also compare our VA results to IV using agricultural prices, wages and weather to instrument for child time-use. This is only feasible for Ethiopia, where these instruments are strong because child labor is prevalent and heavily agricultural. The IV results are similar to VA in that work and chores both have similar negative effects relative to school/study. But they differ from VA in that they imply larger negative effects of both work and chores relative to leisure. Under a LATE interpretation, this may be because work/chores are particularly detrimental for "marginal" children pushed in and out of child labor by the variation of the instruments.

We proceed as follows: Section II reviews the literature, Section III presents a simple theory model to guide the analysis. Section IV describes the Young Lives data and Section V describes our econometric methods. Sections VI to VIII present results and IX 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 children in the world

---

[5] These studies rely on simulation (Guarino, Reckase, and Wooldridge 2014) or comparison of experimental and VA estimates (Angrist, Pathak, and Walters 2013; Deming et al. 2014; Muralidharan and Venkatesh 2013).

today (or 152 million) participate in what they define as child labor.[6] Prevalence is highest in Africa where 20% of children are engaged in child labor, most of which is agricultural. The numbers are higher if household chores are 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 2017a).

There is broad consensus in policy circles that child labor is detrimental for child development, making its elimination a high-profile issue in Low to Middle Income Countries (LMIC). The UN Sustainable Development Goals call for the eradication of all forms of child labor by 2025. The great majority of the world's children live in countries that have ratified the ILO's two main child labor conventions (ILO 2017b),[7] and, according to the ILO, from 2004 to 2014, 57 LMICs implemented 279 policies, plans and programs aimed at reducing child labor. There has also been growing international pressure on companies to avoid using child labor. An example is the debate in the US over the Child Labor Deterrence Act.[8]

Underlying the broad policy consensus for the eradication of child labor is the presumption that working is harmful for children. Given the widespread acceptance of this view, it is perhaps surprising that the academic literature on the effect of child labor on child development has not reached definitive conclusions. There are several sources of ambiguity:

Much of the evidence on child labor concerns its effects on schooling. Most studies find negative effects of child labor on schooling (see Assad et al. 2010, Beegle et al 2009, Boozer and Suri, 2001, Buonomo Zabaleta 2011, Sedlacek et al. 2009), although some find complementarity (Patrinos and Psacharopoulos 1997; Ravallion and Wodon 2000), with child labor up to a certain level contributing to schooling (Ray and Lancaster 2003).

The weakness of schooling as a measure of human capital is well understood. Glewwe (2002) finds direct skill measures predict wages better than schooling. Child labor may hinder learning even if it does not reduce school enrolment (Gunnarsson et al., 2006, Dumas, 2012). Conversely, non-school activities, including child labor, may generate human capital. This suggests focusing on direct measures of human capital, like ability test scores.

---

[6] The ILO defines child labor as "work that is harmful to children's physical and mental development which includes work that … 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).

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

[8] This bill would disallow import of goods produced using hazardous forms of child labor. It has spurred debate about costs and benefits of banning child labor, but it has never passed. It was not until 2015 that import of goods made by forced labor, including forced or indentured child labor, was officially banned in the Trade Facilitation and Trade Enforcement Act of 2015.

However, evidence on the impact of child labor on direct human capital measures is limited. Early studies found negative correlations between child labor and ability test scores (Akabayashi and Psacharopoulos 1999; Heady 2003). Recent studies try to establish causality using panel data or IV methods, and most find negative effects (e.g., Gunnarsson et al. 2006; Bezerra et al. 2009). The work by Emerson et al. (2017) may be the clearest evidence to date of a negative effect: Using fixed effects models to control for latent child ability, they find work lowers test scores by an amount equivalent to the effect of a quarter to three fifths of a year of school. In contrast, using controls for lagged scores, Dumas (2012) finds that, in Senegal, child work actually has (small) positive effects on math scores.

There are important limitations to existing studies based on test scores. The studies we cite above only have data on test scores for the subsample of children who are ***at school***. This may exclude children most adversely affected by child labor (i.e., those who drop out). Also, the use of panel methods to study evolution of test scores requires tests that are comparable over time (Das and Zajonc 2010), and such comparability is rarely achieved in existing data. Two key advantages of the Young Lives data are that it gathers test scores for *all* children regardless of whether they are in school, and it provides tests that are comparable over time.

As we stressed in the introduction, an important limitation of existing studies is failure to account for the complete time budget of children.[9, 10] This is inconsistent with evidence that the trade-off between work and school hours is not one-to-one. For instance, using data from UNICEF's MICS project, Edmonds (2009) finds work hours can increase up to a threshold with little effect on school attendance. Similarly, conditional cash transfer programs that incentivize school attendance tend to reduce child labor by less than they increase school hours, suggesting that substitution away from leisure is also at play (Attanasio et al. 2010; Ravallion and Wodon 2000). In general, the impact of child work on cognitive development will depend on how it affects the whole time budget of children, not just time in school.[11]

---

[9] A notable exception is Akabayashi and Psacharopoulos 1999, who look at effects of school attendance and time spent working and studying on reading and math ability. But a key limitation of their study is that the math and reading measures are simply parental reports of whether the child could read or do math.

[10] Some theory models consider more complete time-use vectors (Cigno and Rosati 2005; Edmonds 2007).

[11] Fiorini and Keane (2014) discuss the importance of accounting for the complete time budget in a developed country setting. We illustrate their point via a simple example: Assume a child's total time (T) is allocated between work (T_{wi}), leisure (T_{li}) and school (T_{si}). Let test scores 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. Now, suppose time not working is split equally between leisure and school, $T_{li} = T_{si} = (T - T_{wi})/2$. Finally, suppose we follow a common practice in the literature and regress the test score Y_i on work time *alone* – without controlling for leisure and school time. In that case, we will obtain a negative estimated effect of child work of $(1/2)(-2) = -1$. This is despite the fact that a negative effect only exists if work crowds out school, not if it crowds out leisure. Furthermore, our result *underestimates* the potential to improve scores by shifting time from work to school, which has a positive effect of 2 units per extra unit of school time (not just 1 unit).

The Young Lives data do attempt to measure the complete time budget of children. This enables us to fill a key a gap in the literature by asking how the complete time budget affects child cognitive development. Also, while many prior studies focus on binary work measures (e.g., Psacharopoulos 1997; Ravallion and Wodon 2000; Emerson et al. 2017), we are able to study how total hours of child work affects child outcomes.

Young Lives separately classifies market work, farm work, chores and work in the family business. This lets us test if different types of work have different effects on child development. Most prior studies focus on market work (Beegle et al 2009; Gunnarsson et al. 2006). But Edmonds (2007) descriptive analysis of MICS data suggests the total time a child spends on all types of work is likely to matter for human capital, not just market activities.

III. A Simple Theoretical Framework

A key challenge in studying effects of child labor is endogeneity of children's time use. To elucidate the problem, consider a simple unitary household model, where a household makes decisions about adult work hours (h), child work hours (h_c) and child school hours (s). The household faces the budget constraint:

$$C = wh + w_c h_c - ps + N$$

where C is consumption, w is the adult wage, w_c is the child wage, p is the price of school, and N is non-labor income. The household also faces the time constraints:

$$l = H - h ; \qquad l_c = H - h_c - s$$

where H is available time, and l and l_c are adult and child leisure, respectively. The final constraint is the production function for child cognitive ability (Y), which takes the form:

$$Y = \mu + s\beta_s - h_c\beta_h$$

Here μ is the child skill endowment, β_s is the effect of s on Y, and β_h is the effect of h_c on Y, both measured relative to leisure (the residual category). The household maximizes a utility function $u(C, Y, l, l_c, s)$. Without loss of generality, we let utility depend on both child leisure and school time, with child work hours the residual category. In this model the three decision variables (h, h_c, s) are driven by four exogenous factors – namely w, w_c, p and N – as well as the child skill endowment μ. An interior solution satisfies the first order conditions:

$$u_c w = u_l \qquad u_c w_c - u_Y(\mu)\beta_h = u_{l_c} \qquad u_Y(\mu)\beta_s + u_s(\mu) - u_c p = u_{l_c}$$

We write $u_Y(\mu)$ and $u_s(\mu)$ to highlight the fact that the child skill endowment μ plausibly shifts both the marginal utility of child cognitive ability and school hours. For example, it is plausible youth with higher μ have a greater marginal utility of school hours $u_{s\mu}>0$ because they are better at school and find it easier (see Keane and Wolpin, 1997).

This simple model illustrates how the five factors w, w_c, p, N and μ influence the solution (h^*, h_c^*, s^*). For clarity, consider a special case with $p=0$, $u_{YY} = 0$ and parent work predetermined. Then, total differentiation of the FOCs reveals that:

$$\frac{ds^*}{d\mu} = -u_{s\mu}/[u_{ss} + u_{l_c l_c} R_c] > 0$$

where $R_c = (u_{cc} w_c^2/[u_{cc} w_c^2 + u_{l_c l_c}])$. That is, school hours are increasing in child ability in the plausible case that $u_{s\mu} > 0$, provided that $u_{ss} < 0$, $u_{l_c l_c} < 0$ and $u_{cc} < 0$. We also have:

$$\frac{ds^*}{dN} = u_{cc} w_c/[u_{ss} + u_{cc} w_c^2 R_l] > 0$$

where $R_l = ([u_{ss} + u_{l_c l_c}]/u_{l_c l_c})$, so wealthier households choose a higher level of child schooling.[12] The main point is that the child time allocation (h_c^*, s^*) is driven by <u>both</u> the exogenous factors (w, w_c, p, N) and the child skill endowment μ.

The upshot is that if we attempt to estimate an empirical version of the production function for child cognitive ability $Y = \mu + s\beta_s - h_c\beta_h$ we face the problem that s and h_c are likely to be correlated with μ. But on the plus side, there is also exogenous variation in s and h_c induced by (w, w_c, p, N). We seek to identify the production function from the latter.

There are three main approaches to this problem. <u>First</u>, we could estimate a structural model, estimating the production function jointly with households' decision rules for child schooling and child work. <u>Second</u>, we could take an IV approach, using exogenous factors (w, w_c, p, N) to instrument for s and h_c. We then identify β_s and β_h from exogenous variation in s and h_c induced by these instruments. Relatedly, one might use experimental designs that shift the constraints so as to generate additional exogenous variation in the inputs. <u>Third</u>, we could control for μ when estimating the production function. If panel data are available, this requires differencing out μ, or using a lagged ability measure to proxy for μ (see Todd and Wolpin, 2003, 2007). In a panel-data approach, we identify β_s and β_h from the variation in s and h_c induced by the variation in exogenous factors (w, w_c, p, N) holding μ fixed.

As we noted in Section II, a few papers have tried to infer causality by instrumenting for child work (Gunnarsson et al. 2006, Bezerra et al. 2009, Dumas, 2012). But the use of IV does not change the fact that it is important to control for children's complete vector of time

---

[12] If we allow for $u_{YY} < 0$ then the sign of $ds^*/d\mu$ becomes ambiguous, as it depends on the relative magnitude of u_{YY} and $u_{s\mu}$. But ds^*/dN remains unambiguously positive. ds^*/dw is also unambiguously positive, while the sign of ds^*/dw_c is ambiguous, as it depends on income and substitution effects. Allowing for $p>0$ and endogenous parent hours greatly complicates the analysis but doesn't change the main points.

uses when estimating the effect of child work. This makes an IV strategy hard to implement, as it requires finding <u>several</u> valid instruments for the <u>vector</u> of endogenous time uses.

In our main analysis we use the panel data value added (VA) approach to control for latent child ability (μ). We will compare this to the panel data fixed effects approach as a robustness check. We also implemented IV using agricultural wages/prices and weather to instrument for children's time use. However, this strategy only generated strong instruments in the case of Ethiopia at younger ages, so we only report IV results for that case.

Prior to our work, only a couple papers used panel data to control for latent ability via lagged human capital measures (Dumas 2012), or fixed effects (Emerson et al. 2017). We build on this work by exploiting the panel aspect of the Young Lives data to estimate value added (VA) and fixed effect models. Young Lives collected – for the first time – test scores that are *both* comparable over time and obtained regardless of school attendance. Prior panel data analysis has not had the benefit of repeated measures with both these features.

Finally, even if child labor hampers child development *ceteris paribus*, it may still be an important source of household income. Child labor bans may harm children if they reduce household resources (Basu 1999; Jacoby and Skoufias 1997). Thus, a complete account of effects of child labor requires controlling for family income and goods inputs into child development. This has not been done in prior work. Young Lives is not perfect in this dimension either, but it does contain measures of parent education and household wealth that we use to proxy for family income and goods inputs.

IV. The Young Lives Data

Young Lives collects information on two cohorts of children from four 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. Thus, the YC is covered from birth roughly through age 12, while the OC is covered from roughly ages 8/9 through ages 19/20.

The sample contains about 12,000 children, consisting of 2,000 YC and 1,000 OC children from each country.[13] Data were collected from 20 sites in each country, selected to represent each country's socio-economic and geographic diversity, with a pro-poor bias.[14]

---

[13] In Peru only 714 OC children were enrolled due to capacity constraints during the first round.
[14] 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.

Then, 100 YC and 50 OC children were randomly sampled from each site. Young Lives collects detailed information about the education, health and development of the target child, as well as detailed data on household economic circumstances and demographics.

All data were collected through home visits by interviewers; these include interviews with the person who knows most about the socio-economic circumstances of the household, and the primary caregiver of the target child. The interviewer also does a direct assessment of the target child. At older ages the target child was also interviewed. In each round, an intense effort was made to find children who had moved since the prior survey. As a result, attrition from 2002 to 2013 is very low – only 3.6% in the YC and 8.3% in the OC.[15]

The 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 the second, third and fourth rounds (2006, 2009, 2013). The YC children were on average age 5, 8 and 12 in rounds 2, 3 and 4, respectively; the OC children were 12, 15 and 19. Age at interview was very similar for children within each cohort/round cell. We control for lagged time allocation and test score variables in our main VA specifications, so the round two data is only used to form these lags. Combining the two cohorts, we can study outcomes at ages 8, 12, 15 and 19. Table 1 shows the total (N) and analysis sample sizes for each country at each of these ages.

Table 1: 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

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 a few cases 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 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.

---

[15] These figures increase to 5.5% and 9.1% if we include mortality. Country specific attrition rates (including mortality) for the YC were 6.3% in Ethiopia, 4.8% in India, 7.3% in Peru and 3.6% in Vietnam, while those for the OC were 9.2% in Ethiopia, 5.6% in India, 11% in Peru and 11% in Vietnam.

Second, time-use data were only collected for at age 5 and older, and not all children had turned 5 at the time of the Round 2 survey. So we lose some of the age 8 observations for the YC. We lose about 10% for India, Peru and Vietnam, but we lose about 20% in Ethiopia.

In the following subsection IV.A we describe the time use data in detail. Then in Subsections VI.B to VI.C we describe the cognitive test data in detail.

IV.A *Children's Time-Use*

A key feature that makes Young Lives uniquely suitable for our analysis is the data on child time use. It is collected starting in round 2, when the Younger Cohort were 5 and Older Cohort were 12. Young Lives does not collect time diaries, but it measures time spent on eight activities in a "typical day"[16] in the last week: (1) sleep; (2) caring for other 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 and other activities outside the household; (6) school (including travel time); (7) studying outside of school (at home, extra tuition); (8) play time/general leisure. For our analysis, we combine categories (2) and (3) into a single category that we call "domestic chores," and we combine categories (4) and (5) into a single category we call "market/farm work."

Table 2: Child Time Allocation – Mean Hours of Each Activity in a "Typical Day"

	Sleeping	Domestic chores*	Market/farm work*	Chores + Market/farm work	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
Vietna							
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

---

[16] 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.

domestic tasks (such as fetching water, firewood, cleaning, cooking and washing); Market/farm work includes 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.

Table 2 presents summary statistics for time allocation. Work at ages 8 and 12 is most prevalent in Ethiopia, where children average 4 hours per day on combined work activities (i.e., market/farm work plus domestic chores). Peru is next with combined hours of 1.63 per day at age 8 and 2.37 at age 12. Work time at young ages is lowest in India, where combined work is 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), and lowest in Ethiopia (about 6 to 7 hours per day). An interesting fact is that school time in Vietnam is lower than in any other country, but home study time is much higher.

 In all four countries, children at ages 8, 12 and 15 spend (on average) more time on domestic chores than market/farm work, typically by a wide margin. But market/farm work become the more common type of work at age 19. Combined work time increases sharply in India, Peru and Vietnam at age 19, catching up to the levels in Ethiopia.

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 3: Prevalence and Intensity of Chores and Work by age and country

	Work Any	Domestic chores (any)	Market/farm work (any)	Domestic chores *and* Market/farm work	Domestic chores hours (if any)	Market/farm work (if any)	Total (if any chores or market/farm work)	N
Ethiopia								
Age 8	0.93	0.85	0.40	0.33	2.89	3.67	4.25	12
Age 12	0.94	0.85	0.48	0.39	2.87	3.41	4.34	18
Age 15	0.99	0.92	0.45	0.38	3.45	3.77	4.92	94
Age 19	0.91	0.72	0.52	0.34	3.71	6.27	6.58	85
India								
Age 8	0.38	0.38	0.01	0.01	1.48	2.39	1.53	16
Age 12	0.69	0.68	0.04	0.03	1.43	4.37	1.65	18
Age 15	0.77	0.72	0.23	0.19	2.33	6.29	4.12	94
Age 19	0.88	0.82	0.41	0.36	3.14	7.67	6.57	92
Peru								
Age 8	0.76	0.75	0.16	0.14	1.84	1.57	2.12	18
Age 12	0.90	0.88	0.25	0.23	2.12	2.02	2.65	18
Age 15	0.91	0.89	0.32	0.30	2.46	3.28	3.56	65
Age 19	0.91	0.74	0.52	0.35	3.61	7.44	7.21	58
Vietnam								
Age 8	0.54	0.52	0.04	0.03	1.53	2.30	1.66	17

Age 12	0.77	0.74	0.21	0.18	1.84	2.06	2.32	18
Age 15	0.93	0.87	0.33	0.26	1.86	4.45	3.29	92
Age 19	0.93	0.80	0.57	0.44	2.28	7.33	6.44	77

Table 3 breaks down work time along the extensive and intensive margins. It reports the proportion of children (by age/country) engaged in any chores or market/farm work, as well as average hours conditional on chores/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 market/farm work in all countries and at all ages. At age 8, the fraction of children engaged in market/farm work ranges from the low single digits in India and Vietnam to 16% in Peru and 40% in Ethiopia. But by age 15 about ¼ to ½ of children are engaged in market/farm work.

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 market/farm work is less common than chores, a typical pattern across ages and countries is that, conditional on engaging in each activity, market/farm work tends to take up more hours than chores. Also, while less than half of children who do domestic chores also engage in market/farm work, the majority of the children who are involved in market/farm work also do chores (consistent with MICS data in Edmonds 2007).

Figure 1 presents descriptive statistics on how school, study and leisure time vary with total hours spent on market/farm work, combining data from all four countries. The horizontal axis reports hours spent on market/farm work, which varies 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 market/farm work 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 market/farm work reaches 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.[17]

Of course, very few 8 year-olds spend 6+ hours per day in market/farm work, but we see the same basic pattern at ages 12, 15 and 19. That is, school hours fall gradually at low levels of market/farm work 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 labor will have a much more negative impact on child cognitive development once work hours reach the level that sharply impacts on school hours.

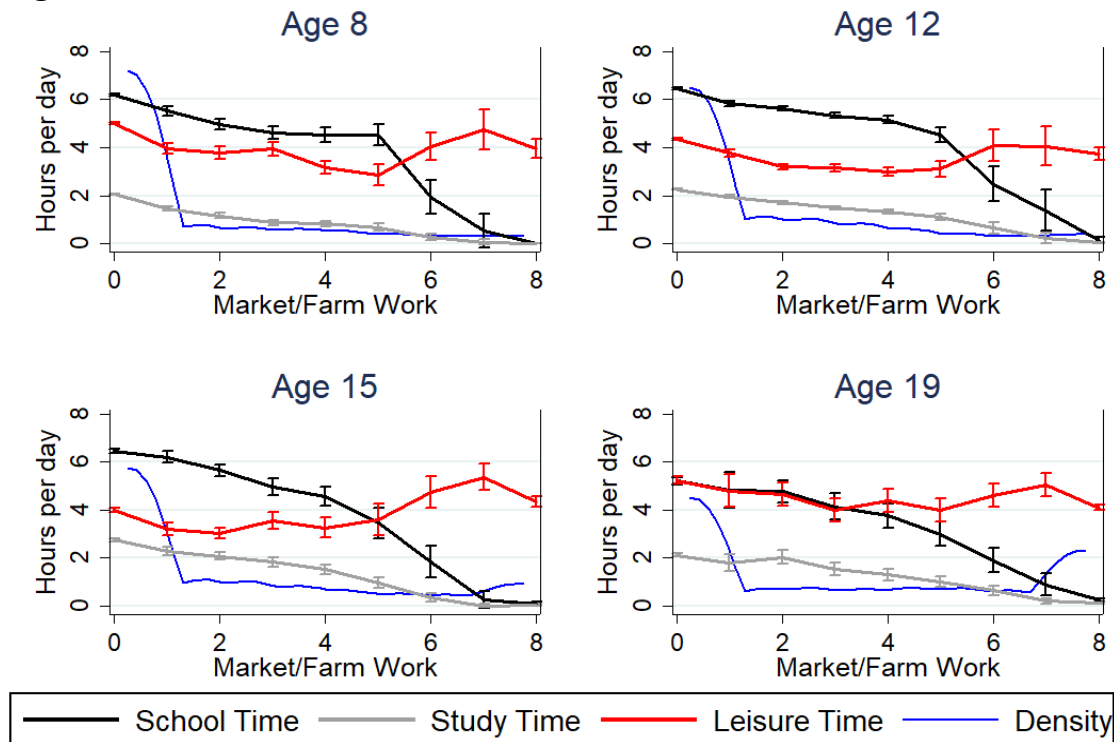**Figure 1: Evolution of School and Leisure Hours with Market/Farm work**

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

---

[17] 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.

Age 8	Age 12
Age 15	Age 19

School Time ——— Study Time ——— Leisure Time ——— Density
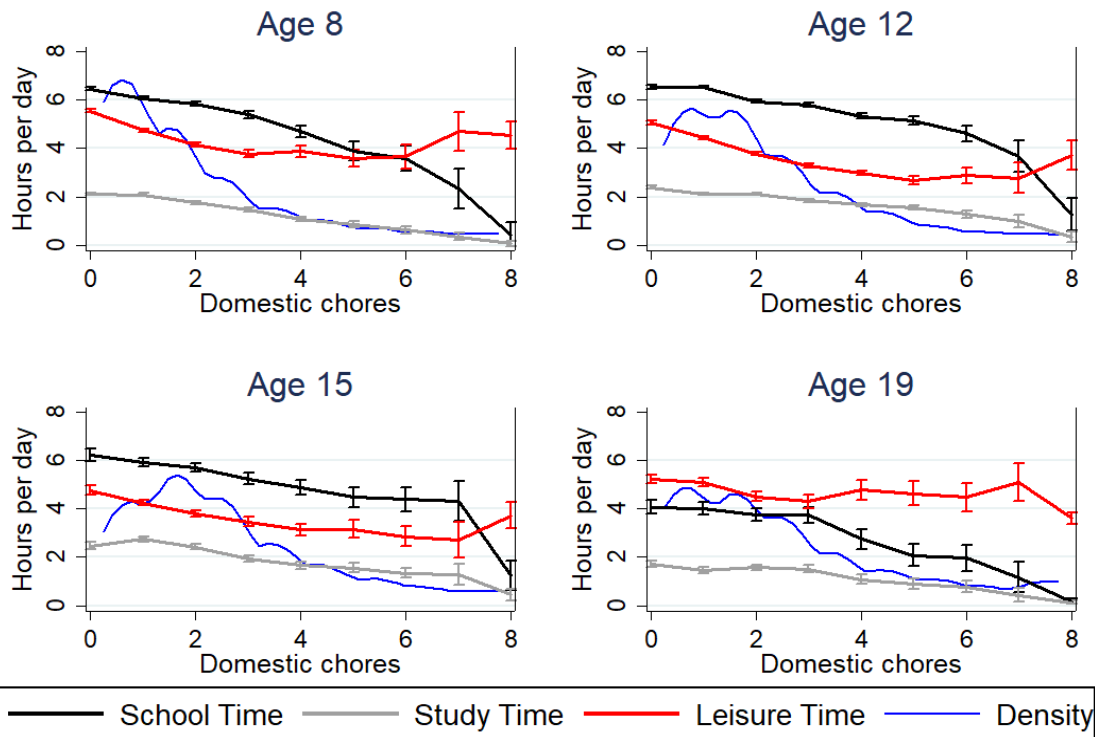
IV.B. *Child Skill Assessments*

The second key feature of Young Lives that makes it ideal for our analysis is the high-quality of the cognitive skill measures. Math and verbal skills were tested regardless of whether a child was in school, allowing us to avoid the selection bias present in prior studies.

Math tests were administered to both cohorts in 2006, 2009 and 2013. So they are available for the younger cohort at ages 5, 8 and 12 and the older cohort at 12, 15 and 19. Math skills at age 5 were assessed by the Cognitive Development Assessment (CDA), designed by the International Association for the Evaluation of Educational Achievement to assess cognitive development of young children (see Montie et al. 2006).[18] After age 5, math skills are assessed using paper-based tests 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.

Verbal skills are measured by the Peabody Picture Vocabulary Test (PPVT).[19] The PPVT was given to both cohorts in 2006 and 2009 and the young cohort in 2013. Thus, it is available for the younger cohort at ages 5, 8 and 12 and the older cohort at ages 12 and 15. In the PPVT, the child is asked to select a picture that best gives the meaning of a stimulus word

---

[18] The CDA 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 fewer cupcakes."

[19] 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).

presented orally by the examiner.[20, 21]

IV.C. *Item Response Theory*

A major challenge in estimating child cognitive ability production functions is the construction of ability measures that are comparable across ages and countries. To achieve this, we use Item Response Theory (IRT) to model the math and verbal test scores. We then construct estimates of *latent* math and verbal ability that are comparable across ages and countries. IRT has a long history in education and psychometrics (e.g., van der Linden and Hambleton 1997). It underlies the construction of well-known tests like the SAT, and it has been used 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.

In the IRT model, each item (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)
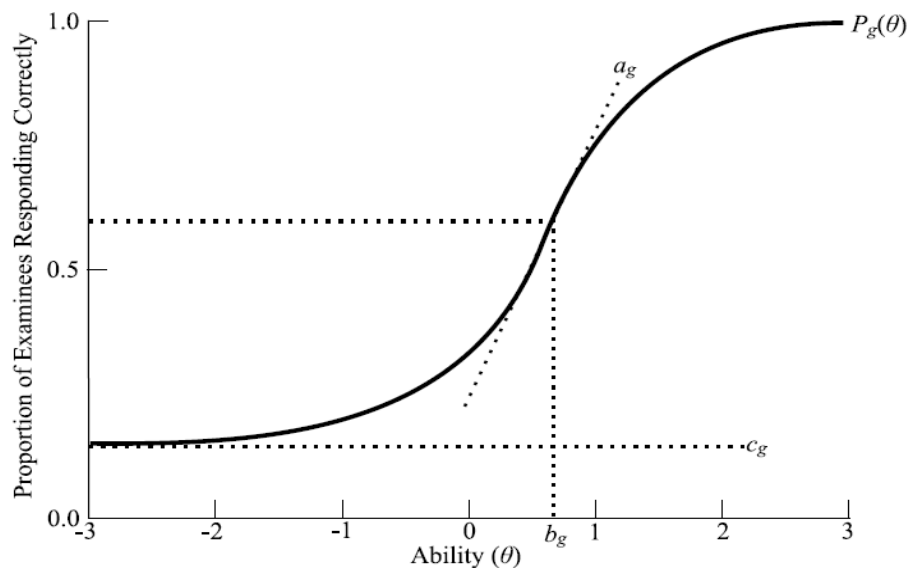
The parameter c_g in Figure 3 is the probability a person can guess the correct answer to item g, given he/she has essentially no ability to actually determine the correct answer.

---

[20] Several studies find a strong positive correlation between PPVT and commonly-used intelligence tests, such as the Wechsler and McCarthy Scales (Campbell 1998; Campbell, Bell, and Keith 2001; Gray et al. 1999).
[21] Official PPVT test manuals include tables for standardisation of the raw scores. But these procedures were not followed in Young Lives as the standardisation samples have different characteristics from the YL samples.

Parameter b_g is known as the "difficulty" of item g. It measures the ability level at which the probability of a correct answer lies half-way between the probability of a correct guess and 100%.[22] Parameter a_g is the "discriminating" power of item g. If a_g is large, the probability a person answers 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 $P_g(\theta)$ can be written as $P_g(\theta|a,b,c) = c_g + (1 - c_g)F[a_g(\theta_i - b_g)]$, where $F(\cdot)$ is the standard normal cumulative distribution function (CDF).

Given responses of a sample of individuals to a set of test questions (with sufficient overlap of questions across people), one can use maximum likelihood to jointly estimate the characteristics (a, b, c) of all items and the ability levels of all individuals. For an *existing* test with a known ICC, one can estimate a person's ability level as the value of θ_i that maximizes the posterior probability of his/her entire set of test responses.

A key advantage of IRT is that, conditional on having partial overlap of test questions across time/context, we can link the ability scores on a common metric. This allows us to construct measures of achievement that are comparable across ages, cohorts and countries. Assessments administered by Young Lives have partial overlap of questions over time and (for math) across countries. Using IRT, the tests can be linked as follows: (i) math tests at age 5 can be linked across countries, (ii) math tests at ages 8-19 can be linked across ages and countries, and (iii) verbal tests (within-language) can be linked across ages 5 to 15. The age 8-19 math scores are normalized to have a mean of zero and a standard deviation of one in the sample of 8-year olds pooled across countries. Similarly, the PPVT scores are normalized with reference to the 5-year old age group within language.

Figure 4: Testing IRT Assumptions

---

[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%.

DIF analysis
Maths pooled

Combining OC and YC from R2 to R4
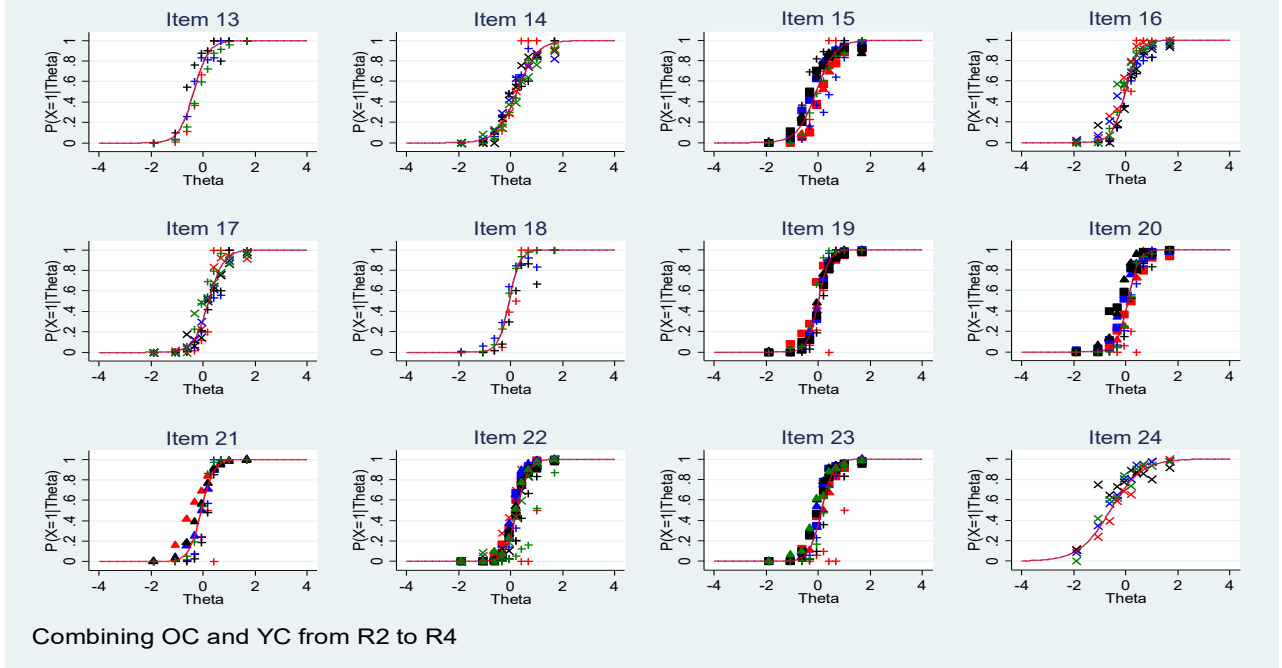
Notes: Ethiopia – Red, India – Blue, Peru – Black, Vietnam – Green
Young Cohort: (R3) +, (R4) ▲, Older Cohort: (R2) ●, (R3) x, (R4) ■
For each item/country/cohort/round we plot the empirical frequency of a correct answer at deciles of ability. If pooling is acceptable then all ability points should lie 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.

Another key advantage of IRT is that it offers a way of running diagnostics on the cross-cultural comparability of test items. Figure 4 presents a test of pooling math test results across the four countries, two cohorts, three rounds and four ages. It plots the 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.[23]

IV.D. *Descriptive Statistics*

Table 1 of the Supplementary Appendix reports summary statistics for round 2 in 2006 when the YC were 5 and the OC were 12. Parental education is much lower in Ethiopia and India than in Peru and Vietnam. About half (44-58%) of mothers in Ethiopia and India

---

[23] 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.

have no schooling, compared to only about a tenth in Peru and Vietnam. Also, compared to other countries, Ethiopian children tend to live in larger households with more brothers and sisters. The age gap between mothers and fathers also tends to be greater in Ethiopia, where mothers are ten years younger than fathers on average.

The Ethiopia, India and Vietnam samples are mostly rural, while Peru is more urban. Over 90% of households in Peru identify as "Mestizo," and in Vietnam over 80% are from the majority Kinh group. Ethiopia and India have more ethnic variation.[24] In all countries, the average child is more than a standard deviation shorter than a typical healthy child of their age (according to the WHO), reflecting the pro-poor bias of the survey.

V. Estimation

To fix ideas, let the production function for *measured* skill Y of child i at age a be:

$$Y_{ia} = F_a(\boldsymbol{X}_i(a), \boldsymbol{U}_i(a), \mu_i, m_{ia}) \tag{1}$$

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 history of observed and unobserved inputs into child development, while μ_i is child i's innate ability (or latent skill), and m_{ia} captures transitory measurement error in the skill test. Assuming for simplicity we observe a child for two periods, and the production function is linear, we have:

$$Y_{i1} = \boldsymbol{x}'_{i,1}\boldsymbol{\beta}_1^1 \qquad\qquad + \boldsymbol{u}'_{i,1}\boldsymbol{\gamma}_1^1 \qquad\qquad + \rho_1\mu_i + m_{i1} \tag{2}$$

$$Y_{i2} = \boldsymbol{x}'_{i,2}\boldsymbol{\beta}_2^2 + \boldsymbol{x}'_{i,1}\boldsymbol{\beta}_1^2 + \boldsymbol{u}'_{i,2}\boldsymbol{\gamma}_2^2 + \boldsymbol{u}'_{i,1}\boldsymbol{\gamma}_1^2 + \rho_2\mu_i + m_{i2} \tag{3}$$

where $\boldsymbol{\beta}_t^a$ is the effect of an age t input on ability at age a, and similarly for $\boldsymbol{\gamma}_t^a$.

A great challenge in estimating this model is omitted variables. We do not observe μ_i or $(\boldsymbol{u}_{i1}, \boldsymbol{u}_{i2})$ as no dataset can contain a complete history of relevant inputs. An endogeneity problem arises if these unobserved inputs are correlated with the observed inputs. We now discuss two ways to deal with this problem using panel-data: First, under the assumption that $\rho_2 = \rho_1 = \rho$, a first-difference (FD) transform will eliminate latent skill, as in:

$$Y_{i2} - Y_{i1} = \boldsymbol{x}'_{i,2}\boldsymbol{\beta}_2^2 - \boldsymbol{x}'_{i,1}(\boldsymbol{\beta}_1^1 - \boldsymbol{\beta}_1^2) + \left(\boldsymbol{u}'_{i,2}\boldsymbol{\gamma}_2^2 + \boldsymbol{u}'_{i,1}(\boldsymbol{\gamma}_1^2 - \boldsymbol{\gamma}_1^1) + m_{i2} - m_{i1}\right) \tag{4}$$

Thus, estimation of a first-difference (or fixed effects if $T>2$) specification gives consistent estimates of effects of *observed* inputs – time-use variables in our case – if: (i) the effects of

---

[24] The main reported religious affiliations are Orthodox Christian in Ethiopia, Hindu in India, Roman Catholic in Peru and "no affiliation" in Vietnam.

unobserved inputs are time invariant (i.e., $\boldsymbol{u}'_{i,2}\boldsymbol{\gamma}^2_2 = \boldsymbol{u}'_{i,1}(\boldsymbol{\gamma}^1_1 - \boldsymbol{\gamma}^2_1)$) so they drop out of (4), or if (ii) observed inputs are uncorrelated with the non-zero part of $\boldsymbol{u}'_{i,2}\boldsymbol{\gamma}^2_2 - \boldsymbol{u}'_{i,1}(\boldsymbol{\gamma}^1_1 - \boldsymbol{\gamma}^2_1)$.

Alternatively, one may use a value added (VA) specification to attempt to control for unobserved ability and other unobserved inputs. The standard VA model has the form:

$$Y_{i2} = \rho Y_{i,1} + \boldsymbol{x}'_{i2}\boldsymbol{\beta}^2_2 + \left(\boldsymbol{u}'_{i,2}\boldsymbol{\gamma}^2_2 + \varepsilon_{i2}\right) \tag{5}$$

where we make explicit the unobserved inputs in the error. Substituting for Y_{i1} we obtain:

$$Y_{i2} = \boldsymbol{x}'_{i2}\boldsymbol{\beta}^2_2 + \boldsymbol{x}'_{i,1}\rho\boldsymbol{\beta}^1_1 + \boldsymbol{u}'_{i,1}\rho\boldsymbol{\gamma}^1_1 + \rho\rho_1\mu_i + \rho m_{i1} + \boldsymbol{u}'_{i,2}\boldsymbol{\gamma}^2_2 + \varepsilon_{i2} \tag{6}$$

Note that (6) is equivalent to (3) if (i) the effect of the latent skill endowment depreciates at the rate ρ, so $\rho_2 = \rho\rho_1$, (ii) the effect of lagged inputs (both observed and unobserved) also depreciate at rate ρ, so that $\boldsymbol{\beta}^2_1 = \rho\boldsymbol{\beta}^1_1$, and $\boldsymbol{\gamma}^2_1 = \rho\boldsymbol{\gamma}^1_1$, and (iii) $\varepsilon_{i2} \equiv m_{i2} - \rho m_{i,1}$. Thus, in a VA model, the lagged test score controls for latent ability and lagged inputs if the common depreciation rate assumption holds.[25]

The error term in the VA model in (5) contains unmeasured current period inputs, $\boldsymbol{u}_{ia}$, as does the error term in the fixed effects (FE) model in (4) if $\boldsymbol{u}'_{i,a}\boldsymbol{\gamma}^a_a \neq \boldsymbol{u}'_{i,a-1}\boldsymbol{\gamma}^a_{a-1}$. In each case, endogeneity arises if $\boldsymbol{u}_{ia}$ is correlated with measured current inputs ($\boldsymbol{x}_{ia}$).[26] Using IV to deal with this problem is difficult, as we have several endogenous time inputs, so we need several instruments. Instead, we include measures of family background and non-time inputs in our VA and FE models in an attempt to control for $\boldsymbol{u}_{ia}$ as well as possible. [27, 28] Letting $\boldsymbol{b}_{ia}$ denote this vector of controls, and extending our model to a periods, we rewrite (5) as:

$$Y_{ia} = \rho Y_{i,a-1} + \boldsymbol{x}'_{ia}\boldsymbol{\beta}^a_a + \boldsymbol{b}'_{ia}\boldsymbol{\delta}_{ia} + \varepsilon_{ia} \tag{7}$$

As Todd and Wolpin (2003) stress, the fixed effects and VA models give consistent estimates of effects of observed inputs on child ability under different conditions. The fixed

---

[25] With more than two periods conditions (i) and (ii) generalize to $\rho_a = \rho^a \; \forall \; a$, and $\boldsymbol{\beta}^a_{a-k} = \rho^k \boldsymbol{\beta} \; \forall \; a$, and $\boldsymbol{\gamma}^a_{a-k} = \rho^k \boldsymbol{\gamma} \; \forall \; a$. That is, the VA model assumes effects of lagged inputs $\boldsymbol{x}_{i,a-k}$ and the initial skill endowment μ_i on Y_{ia} all depreciate at rate ρ. Of course, as $\rho \to 1$ the VA and first difference (or FE) models are equivalent.

[26] For example, suppose the nutrition input into child development is unmeasured. A positive shock to the wage of child labor 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. Thus, we may underestimate the *ceteris paribus* negative effect of child work because it coincides with improved nutrition.

[27] Formally, the VA model gives consistent estimates of the effects of *current* observed inputs if (i) and (ii) hold, $Y_{i,a-1}$ is predetermined, and observed current inputs x_{ia} are uncorrelated with unobserved current inputs u_{ia}.

[28] Measurement error in the lagged test score is also a potential problem for the VA model. The error term ε_{ia} in (8) includes $-\rho m_{i,a-1}$. So ε_{ia} is negatively correlated with measurement error in the lagged test score. This biases $\hat{\rho}$ downward, and biases $\hat{\boldsymbol{\beta}}$ in an ambiguous direction. We find instrumenting for the lagged score using a different test shifts $\hat{\rho}$ upward by 50% or more, but has little impact on estimated effects of the time inputs.

effects model assumes latent child ability and other unobserved inputs have constant effects over time, so differencing or de-meaning eliminates them, while the VA approach assumes their effects depreciate at a common rate over time, so lagged test scores control for them. We adopt VA as our preferred approach, as specification tests in Todd and Wolpin (2007) and Fiorini and Keane (2011) favor it, as do the validation studies noted in the introduction. But we also estimate fixed effects models as a robustness check on all our results.

Finally, we present our estimating equation. Following Todd and Wolpin (2007), we relax the VA model's common depreciation rate assumption by estimating an "extended" VA model that adds lagged inputs $x_{i,a-1}$ to (7). We also extend the VA model by allowing the depreciation rate of the lagged score and lagged inputs to vary by age. Then we have:

$$Y_{ia} = \rho_a Y_{i,a-1} + x'_{ia}\boldsymbol{\beta_a^a} + x'_{i,a-1}\boldsymbol{\pi}_a + b'_{ia}\boldsymbol{\delta}_{ia} + \varepsilon_{ia} \qquad (8)$$

We estimate equation (8) separately for children of each age in each country.

In our specification of (8), the time-varying inputs x_{ia} are the vector of child time use variables: time spent on market/farm work, domestic chores, school, study, play/leisure and sleep. The control variables b_{ia} are characteristics of the <u>child</u> (age, gender, religion, ethnicity, height-for-age z-score), the <u>parents</u> (father's and mother's age and education, whether both parents are present), and the <u>household</u> (urban/rural, wealth,[29] siblings and other adults present).[30] Note that the elements of b_{ia} either vary deterministically over time (age), change little over time, or are time-invariant. Thus, we do not include $b_{i,a-1}$ in (8). The full set of control variables included in the main models are listed in Table 1 of the Supplementary Appendix.

VI. Effects of Child Work on Child Development – Disaggregate Results

We present our results in three parts. First, in this section, we estimate the impact of child work on child development at the country/age level. We also compare results using our approach that controls for the complete time budget with "status quo" approaches that do not. Then, in Section VII, we check the robustness of our results to the estimation method (value added vs. fixed effects vs. instrumental variables). Finally, in Section VIII, we focus on heterogeneity in the effect of child labor across countries, age and gender of children.

VI.A. *Child Work and Child Development: Results by Country/Age*

---

[29] We use the wealth index constructed by the Young Lives team. See Note in Supp. Appendix Table 1.
[30] Given the volatility of households' current income, the education and wealth measures are likely to be better proxies than income for goods inputs (assuming goods inputs are largely determined by permanent income).

First, we estimate equation (8) separately by country/age, using math or verbal scores as the dependent variable. For each of the four countries we estimate math equations for ages 8, 12, 15 and 19, and verbal equations for ages 8, 12 and 15. This gives 28 regressions (16 for math, 12 for verbal), so we cannot present all estimates. Instead, Figures 5 and 6 summarize our main results *graphically* by plotting the coefficients on the key variables of interest: hours of domestic chores and market/farm work.[31] The figures also report 95% confidence intervals around the point estimates. We focus on effects of current time use (see the Supplementary Appendix Section 2 for lagged score and time input coefficients).

To identify the model we must choose an omitted time use category.[32] All parameters are then interpreted as the effect of child work *relative* to that category. While the choice of omitted category has no substantive importance, it has practical importance for the proper interpretation of the estimates. For example, if the omitted time category is leisure, the coefficients on domestic chores are interpreted as the effect of increasing time spent on domestic chores by one hour while reducing leisure by one hour.

A simple way to visualize the effect of child work relative to each of the alternative time-use categories is to omit each alternative time-use category in turn, and plot the coefficients on domestic chores and market/farm work in each case. Thus, we will present the effects of domestic chores and market/farm work first relative to leisure, then relative to time in school, and then relative to time spent studying.

We will abstain from commenting on age or country-specific coefficients, and instead focus on overall patterns. The large number of effects we estimate creates a high probability of Type I error, so we do not want to read too much into significance of individual estimates. We defer analysis of age and country-specific effects to Section VIII below.

We begin by looking at simple models where time use enters linearly. Thus, the effects that we estimate here should be interpreted as effects at the margin for a typical child. Later, in Section VIII.D, we study whether effects of child work differ by the level of work.

The first set of results is presented in Figures 5 and 6. These show some striking differences in the trade-offs between child work (domestic chores and market/farm work) and alternative time-uses. We find no clear evidence to suggest that time spent working is less productive for children's math and verbal ability than time spent in leisure. In contrast, there is strong evidence that child work is detrimental for the development of math and verbal

---

[31] The complete set of estimates is available by request.
[32] As we allocate all 24 hours of the day to the possible activities, the five time-inputs are perfectly collinear.

skills if it crowds out school time or study time.

For instance, In Figure 5 Panel A, where leisure is the omitted category, we find that, for math scores, the effect of an extra hour spent on domestic chores instead of leisure is significantly negative (5% level) in only one of the 16 cases. Similarly, only 2 out of the 16 estimates of the effect of time spent on market/farm work relative to leisure are significantly negative.[33] Almost all coefficients are small in magnitude, and 15 of the 32 are the "wrong" sign (i.e., the point estimate of the child work effect is actually positive).

Figure 6 Panel A shows similar results for verbal development. The effect of chores relative to leisure is significantly negative (at the 10% level) in only 1 of the 12 cases, and in 6 cases the estimates are the "wrong" sign. Similarly, only 1 of 12 estimates of effects of market/farm work relative to leisure is significantly negative at the 5% level. Thus, our results so far provide no evidence for the proposition that child work is detrimental for cognitive development, provided that work only crowds out leisure time.

In contrast, we find strong evidence that child work is detrimental for development of math and verbal skills if it crowds out school time. In Figure 5 Panel B, where school is the omitted category, we find that, for math scores, 12 of the 16 estimates of the effect of an extra hour spent on market/farm work instead of school are significantly negative at the 1% level. Similarly, for verbal skills (see Figure 6B), all 12 point estimates of the effect of market/farm work relative to school are negative, and 7 are significant (5 at the 1% level).

Our results for the effects of domestic chores relative to school are somewhat less statistically significant, especially for verbal ability. Nevertheless, all point estimates for math and all but one for verbal ability continue to be negative (see Figure 6, Panel B)

Finally, in Panel C of Figures 5 and 6 we show results with study time as the omitted category. Here there is a clear pattern of mostly significant negative effects of both chores and market/farm work on both math and verbal scores. In fact, our estimates of the negative effects of work time relative to study time are more consistently significant than were our estimates for work relative to school. This suggests that study is more productive for child development than school time *per se* in the LMIC context.

This finding is not surprising given the evidence of exceptionally low school quality in many LMIC's (see Glewwe and Muralidharan, 2016). It is further confirmed by estimates presented in Figure 7, where we make study the omitted category so we can focus on the

---

[33] 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.

effect of school relative to study. In the large majority of cases the point estimates imply time spent studying is more productive than time spent in school. Combining the math and verbal results, we find that in 5 out of 28 cases the difference is significant at the 5% or 1% level.

To put the effect sizes in Figures 5 and 6 in some context, the median point estimate implies an extra hour of market/farm work relative to time at school reduces math scores by 0.04 standard deviations, and verbal scores by 0.065 standard deviations. Recall from Table 3 that the average child engaged in market/farm work in Ethiopia works about 3.5 hours per day at ages 8, 12 and 15. The median 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 average gain in math scores between age 8 and 12, and 25% of the average gain between age 15 and 19. This effect is magnified by the dynamics of the model.

It is interesting that our estimates of the negative effects of domestic chores on child outcomes are only slightly smaller than those for market/farm work. Some prior work finds market/farm work has a more adverse impact (Bezerra et al. 2009; Buonomo Zabaleta 2011; Emerson et al. 2017; Gunnarsson et al. 2006; Orazem and Gunnarsson 2004). In Figure 8 we take a closer look at this issue by making domestic chores the omitted time-use category and focusing on the relative impact of market/farm work. We find little evidence that market/farm work is worse for development than chores. When math score is the outcome, only 2 of the 16 coefficients on market/farm work are significantly negative (5% level), and the remaining 14 are rather tightly clustered near zero. Similarly, for verbal skills, the estimated effect of market/farm work relative to chores is negative and significant in only 3 cases. Our finding that chores have a similarly adverse effect on child development as market/farm work – if they both substitute for educational activities – is at odds with several studies cited above.

VI.B. *Comparison to "Status Quo" Results*

A key feature of our paper is we control for the full time budget of children. Thus, we estimate effects of time spent working *relative* to specific alternative activities (e.g., leisure, school or studying at home). In contrast, the typical approach in the child labor literature is to estimate the effect of work time without controlling for other time uses.[34] We call this the "status quo" approach. In this Section we compare our results to those obtained by applying a status quo approach, in order to assess whether our approach yields new insights.

We replicate the "status quo" type of analysis by estimating the model in equation (8)

---

[34] Examples of this type of analysis are Akabayashi and Psacharopoulos 1999; Bezerra et al. 2009; Buonomo Zabaleta 2011; Heady 2003; Ray and Lancaster 2003.

but including only measures of time spent on domestic chores and market/farm work, rather than all types of time-use. The results are reported in Figure 9. These imply that hours spent on market/farm work have an unambiguous negative effect on math and verbal skills. In contrast, there is no evidence that hours spent on domestic chores have a detrimental effect on verbal skills. There is some evidence that hours spent on domestic chores are detrimental for math skills (i.e., 6 of 16 estimated effects are significantly negative), but the point estimates are generally much smaller in magnitude than we see for market/farm work.

Another group of "status quo" studies include an indicator for whether a child works, rather than actual work hours.[35] We replicate this type of analysis by re-estimating equation (8) including only indicators of whether any time is spent on market/farm work or domestic chores, without controls for other time uses. The results are presented in Figure 10. These results provide no evidence that domestic chores have a detrimental effect on either math or verbal scores. In contrast, they imply that market/farm work is clearly detrimental for both.

Thus, application of the "status quo" analysis to our data yields the conclusion, similar to that drawn in several previous studies, that market/farm work is harmful for child development, while time spent on domestic chores is either not harmful, or much less so. In contrast, our results in Figures 5 and 6 imply that *neither* time spent on domestic chores nor market/farm work is harmful for child development if they substitute for leisure time. But time spent on domestic chores and market/farm work are *both* harmful for development – and roughly equally so – if they substitute for school time or study time.

The fundamental difference in the main take-away from the "status quo" approach relative to our approach can be explained by differences in the allocation of non-work time among children engaged in market/farm work vs. domestic 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 market/farm work is associated with greater reductions in school time than 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 market/farm work 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

---

[35] Examples of this type of analysis are Amin, Quayes, and Rives 2006; Beegle et al. 2009; Emerson et al. 2017; Patrinos and Psacharopoulos 1997; Ravallion and Wodon 2000.

tends to be is associated with larger reductions in school time.

VII. Robustness Checks – Fixed Effects and Instrumental Variable Results

Here we examine the robustness of our results in Section VII.A to alternative methods of controlling for latent child ability and/or unobserved inputs. We start by checking the extent to which unobservables play a role in our results. In our main model the key control for unobserved ability is lagged test score. We also include 1) a lagged time-allocation vector in order to allow the effect of lagged inputs to vary by age, 2) parental education and household wealth to proxy for household resources, and 3) several additional controls that serve as further proxies for past inputs and child ability.[36] At the extreme, if child ability and unobserved inputs play no role in driving the relationship between current time inputs and child development, we would expect to get similar results with and without these controls.

Table 4 reports estimates of associations between math scores and child work from three models. The first ("no controls") includes only the current time allocation vector (study is the omitted category). The second ("background controls") adds the vector of control variables $\boldsymbol{b}_{ia}$. The third ("full model") adds lagged test score and lagged time-use variables.

The "no controls" column shows that for all ages and countries there is a significant and quantitatively large negative correlation between math scores and time spent on both chores and market/farm work (relative to study time). The median point estimate is -.143 for chores and -.137 for market/farm work. Across all ages and countries, controlling for $\boldsymbol{b}_{ia}$ causes the estimated impacts of child work relative to study to drop by roughly 1/3 to one half. The median point estimate drops to -.081 for chores and -.102 for market/farm work. But the estimates remain highly significantly negative in almost all cases. Adding in lagged test scores and time-use causes the median point estimates to drop further to -.045 for chores and -.053 for market/farm work. Nevertheless, the estimates remain highly significantly negative in 11 and 14 (out of 16) cases for chores and market/farm work respectively.

This exercise suggests that the raw correlations between child work and math scores exaggerate the negative impact of work relative to study by a factor of 2 to 3. This is the bias we expect if higher latent ability children study more and/or do less work. But even with extensive controls we still find significant negative effects of child work relative to study for most countries/ages. The results for verbal scores (available on request) are very similar.

---

[36] As described in Section IV these include child age, gender, religion, ethnicity, height-for-age; parental age and whether in household; whether household is urban/rural, household member composition.

VII.A. Fixed Effects Results

Having found that controls for latent child ability are important, we next test how sensitive our estimates are to the specific assumptions made about how unobserved ability and inputs affect child outcomes. To do this we compare our main VA estimates to those from a Fixed Effects (FE) model. The FE model is obtained by generalizing equation (4) to a periods, and assuming our control variables $\boldsymbol{b}_{ia}$ capture the influence of unobserved inputs:

$$Y_{ia} - Y_{i,a-1} = \boldsymbol{x}'_{i,a}\boldsymbol{\beta}^a_a - \boldsymbol{x}'_{i,1}\boldsymbol{\beta}^{a*}_1 + \boldsymbol{b}'_{ia}\boldsymbol{\delta}_{ia} + \varepsilon^*_{ia} \tag{9}$$

where $\boldsymbol{\beta}^{a*}_1 \equiv (\boldsymbol{\beta}^1_1 - \boldsymbol{\beta}^2_1)$ and $\varepsilon^*_{ia} \equiv (m_{i2} - m_{i1})$. This is analogous to how we derived our VA estimating equation (8) from the model in (5). While the VA model assumes unobserved ability is captured by the lagged outcome measure, in the FE model the effect of latent ability is differenced out under the assumption that its effect on child outcomes is age invariant. Large differences between the VA and FE estimates would suggest our results are vulnerable to the specific way in which we control for unobserved ability.

Figure 11 reports the fixed effects results for math. The FE model is somewhat less efficient than value added, for the usual reason that it relies purely on within subject variation (differencing out all between subject variation). For this reason, we merge the school and study time use categories into a single "school/study" category. Also, for math we cannot obtain fixed effects estimates for age 8 because the test at age 5 is not exactly comparable.

As before, we find no evidence that chores or market/farm work are detrimental for child math skills if they substitute for leisure time (Figure 11, Panel A). However, both chores and market/farm work do have detrimental effects on math if they substitute for school/study time (Figure 11, Panel B). Thus, our FE results for math are qualitatively similar to the VA results. The FE results for verbal scores (available on request) are also very similar.

It is comforting that our main results are not sensitive to the specific way we control for latent ability and unobserved inputs. The similarity of the VA and FE results, despite the fact that these estimators give consistent estimates under somewhat different assumptions, gives us added confidence in the robustness of our main conclusions.

VII.B. Instrumental Variables Results

In this section we report instrumental variable results. We are aware of only three prior studies that attempt to infer the effect of child work on cognitive test scores using IV. Gunnarsson et al. (2009) instrument for child work using mandatory school entry/exit ages across nine Latin American countries, along with their interactions with age, gender and

urban/rural residence. Bezerra et al (2009) use the unskilled wage in the state of residence in Brazil. And Dumas (2012) uses rainfall (children work less in droughts) and commuting time to school in Senegal. The results are contradictory, as the former two papers find negative effects of child work, while the latter finds positive effects. All three studies find 2SLS estimates of effects of child work on cognitive test scores that are, perhaps surprisingly, several times *greater* in magnitude than OLS estimates. But the 2SLS estimates are also much more imprecise that the OLS estimates, suggesting weak instruments may be an issue.

Our challenge in finding appropriate instruments is greater, as we seek to infer the effect of a child's entire vector of time use, so we need to instrument for several endogenous variables.[37] To somewhat reduce the problem, we combine school and study into one category (which we treat as the omitted category). We then need to instrument for four endogenous variables: Hours spent on market/farm work, chores, leisure and sleep.

As a guide to finding good instruments, we also examined papers that instrument for child work to estimate its effect on schooling: Boozer and Suri (2001) use regional rainfall in Ghana, while Beegle et al. (2009) use the price of rice at the community level in Vietnam. In the Nicaraguan context, Buonomo Zabaleta (2011) uses a large set of instruments including weather shocks, wage rates, distance to school, and access to electricity and water.[38] We also consulted reduced form studies of how shocks affect child labor: Beegle et al (2006) find that crop loss shocks increase child work and chores, and reduce school attendance in Tanzania. Trinh et al. (2020) find that below average rainfal reduces child work and chores in Vietnam.

Based on this prior work, we decided to construct instruments based on agricultural prices, local wages and prices, rainfall and weather shocks, and distance to school. More specifically: For each household/year we construct an agricultural price index where specific crop price indices are weighted by the quantity of each crop the household grows. We also construct a livestock output price index based on the price of each animal's output (e.g., price of eggs for chickens), weighted by the quantity of each type of animal the household holds.

We construct community wage indices by taking the local wage rates for men and women for a range of occupations reported in Young Lives, and taking the first two principle components of those series. Similarly, we construct an index of consumer prices by taking local prices for a range of common consumer goods reported in Young Lives, and taking the

---

[37] Notably, Dumas (2012) attempted to estimate the effect of both work hours and school, but was not successful in finding strong instruments for school.

[38] Buonomo Zabaleta (2011) also uses school quality, household assets, household head education, household size and number of young children. We chose not to use these sort of variables as excluded instruments as we suspect they may be correlated with household resources, or enter the test score production function directly.

first two principle components of those series.

Next we construct rainfall in the 2 and 12 months prior to the interview using satellite data (see Appendix for details), taking deviations from 20-year means to construct rainfall shocks. These are interacted with interview month to let rainfall have different effects at different times of year. We also use frost shock indicators from the Young Lives data.

Additional instruments are distance to school and whether the nearest school is public. We also construct an index of other local services/infrastructure, including such things as community access to electricity, drinkable water and children's playgrounds. We assumed this would be be useful in predicting time devoted to chores and leisure.

Our instruments can be classified into those that vary at the household level – namely the agricultural and livestock price indices, school distance and type, rainfall in the months prior to the interview date, frost shocks, interview month and whether the household is agricultural – and those that are at the community level – local wage and consumer price indices, and the community service/infrastructure index.

We interacted the community level instruments and rainfall with several child and household chacteristics to let them differentially affect different types of children/households. These are child gender, number of older siblings, number of brothers/sisters, time to school, household wealth, urban/rural, the agricultural and livestock price indices, frost shocks and month of interview. With interactions included, we have about 120 to 130 instruments. This varies slightly by country/age (e.g., there are fewer interview months in some cases).

First we assess whether our instruments are sufficiently strong to make 2SLS feasible, using first-stage F-statistics for weak instruments. We have four endogenous variables, so we use the Sanderson and Windmeijer (2016) F-statistic designed to assess *independent* variation in each endogenous variable generated by the instruments. Given the large instrument set, the many instrument problem may be important, so we consider both 2SLS on the full instrument set, and results using LASSO to select a subset of strongest instruments (see details below).

The results show our instruments are quite strong for Ethiopia at ages 8 and 12. Other cases are borderline at best, so we do not consider them further. Our instruments are strong for Ethiopia at ages 8 and 12 because almost all child labor in that context is agricultural, and our instruments are specifically designed to predict work/chores in an agricultural context. In addition, as we showed in Section IV.A, child labor at young ages is much more common in Ethiopia than the other countries, so we are not trying to predict a rare event.

We report the IV results for age 8 and 12 math scores in Ethiopia in the top panel of Table 5. The first three columns report results from VA and OLS (with and without controls)

27

for comparision purposes. The next three columns report 2SLS using the full instrument set, followed by 2SLS and optimal (2-step) GMM using the subset of instruments chosen by LASSO. The last three columns also report IV results, except here we *both* instrument for time use *and* control for the lagged test score, thus combining the IV and VA approaches.

For Ethiopia at age 8, if we implement 2SLS using the full set of 123 instruments, the coefficients on market/farm work and chores are -.084 and -.090 respectively. This implies an extra hour of work or chores that detracts from school/study time reduces math scores by roughly 0.085 standard deviations. The standard errors are 0.011 and 0.15, respectively, so the point estimates are precise despite instrumenting. But the SW first-stage F-statistics for work and chores are 6.72 and 4.54, respectively. These values raise a concern that the large instrument set may be causing a many/weak instrument problem.[39]

To address this concern, we reduce the instrument set by using LASSO regression to select the best instruments. We use the Bayes information criterion (BIC) to assess the fit of alternative first-stage models, adjusting for number of parameters.[40] To ensure LASSO does not choose instruments that merely give the best fit to each endogenous variable *individually*, but rather to maximize their *independent* variation, we use a multi-step approach: First use LASSO to chose a subset of instruments that best explain the child work variable. Second, use LASSO to choose a subset of instruments that best explain the *residual* variation in the chores variable (not explained by instruments selected in step 1). Repeat for lesiure and sleep.

Adopting this procedure, we chose a subset of 24 instruments, listed in the Supplementary Appendix. Using these 24 as the excluded instruments in 2SLS, the SW F-statistics for work and chores improve markedly to 12.6 and 9.9, respectively.[41] The point estimates for work and chores become slightly more negative, and standard errors increase by only 30%, so the estimates remain precise. In contrast to the IV papers in the prior literature (see above), our instruments give much more precise estimates of child labor effects.

The next column of Table 5 reports optimal (2-step) GMM-LASSO estimates which are very similar to the 2SLS-LASSO estimates. The last three rows of Table 5 report the same models except now we add the lagged test score as an additional control, thus combining the IV and VA approaches to dealing with latent child ability. Controlling for the

---

[39] Stock-Yogo critical values for weak instrument F-statistics in the case of multiple endogenous variables and instrument sets larger than 100 are not available. A rule-of-thumb is we would like to see F-stats of 10 or more.
[40] Of course, we force LASSO to to retain all 2nd stage control variables as instruments in the first stage. LASSO then choses *additional* variables from the set of *excluded* instruments.
[41] With four endogenous variables and 24 instruments, the Stock-Yogo critical value for rejecting the null that IV bias relative to OLS is greater than 10% is 11.4. The Stock-Yogo critical values for rejecting the null of IV bias greater than 30%, 20%, 10% or 5% of the OLS bias are 4.4, 6.2, 11.4 and 21.4 respectively.

lagged test has little impact on the IV results.

The GMM estimates for Ethiopia at age 8, that control for lagged test score and use the 20 instruments chosen by LASSO, give coefficients on market/farm work and chores of -0.091 and -0.116, and these are precisely estimated (the standard errors are .015 and .022, respectively). In one way these results are consistent with our value added (VA) results, in that effects of work and chores are not significantly different. Substituting an extra hour of either for school/study reduces math test scores by roughly <u>a tenth of standard deviation</u>.

However, the IV estimates are roughly twice as large as the VA estimates. There are two plausible explanations: The first is measurement error in the time use variables (which would bias estimated effects of work/chores towards zero). But we are skeptical of this story, because if we use child reports of time use to instrument for parent reports it has little impact on our VA results.[42] The second plausible explanation is based on a LATE interpretation of IV: It is possible the children at the margin – who are shifted in and out of work/chores by the variation in the instruments – are relatively more adversely affected by child labor.

Note that while the IV estimates exceed VA, they are smaller than what we obtain in a simple OLS regression of test scores on the time use vector (with no controls). Thus, we still find it is relatively lower skill children who are selected into work (biasing OLS downward).

The bottom panel of Table 5 reports IV results for math scores at age 12 in Ethiopia. The results closely mirror those for age 8. Our instruments are somewhat stronger here, as the SW F-statistics for those selected by LASSO exceed 14.5 for chores and 15.7 for work. The GMM estimates that control for the lagged test score give coefficients on market/farm work and chores of -0.111 and -0.097, which are similar to what we found at age 8. They again imply that substituting an extra hour of either work or chores for school/study reduces math test scores by roughly <u>a tenth of standard deviation</u>. This is consistent with our value added (VA) results, in the sense that effects of work and chores are not significantly different.

Table 6 reports IV results for verbal scores in Ethiopia at ages 8 and 12. The GMM estimates in the last column – that control for lagged test score and use 20 instruments chosen by LASSO – give coefficients on market/farm work and chores of -0.245 and -0.179 at age 8, with standard errors of .032 and .061, respectively. We cannot reject that these effects are equal, so it appears that substituting an extra hour of either work or chores for school/study reduces math test scores by roughly <u>two tenths of a standard deviation</u>. This is twice as large as the effect we found for math scores. Again, we find the IV estimates are roughly twice as

---

[42] Child reports of their own time use are available at ages 12 and 15 only.

large as the VA estimates. The age 12 results are similar, but the effect sizes are a bit smaller.

Finally, Tables 6 and 7 also report the coefficients on leisure. Our VA results imply that child work only has a negative impact if it substitutes for school/study (not leisure). Our IV results for verbal scores are consistent with this hypothesis. But our IV results for math imply child work has adverse effects relative to leisure in Ethiopia at both ages 8 and 12.

Overall, the IV results are similar to VA in that work and chores have similar negative effects relative to school/study. But they differ from VA in that they imply larger negative effects of both work and chores – both relative to school/study and, for math, relative to leisure. Under a LATE interpretation, this may be because work/chores are particularly detrimental for "marginal" children pushed in and out of child labor by the variation of the instruments. Nevertheless, the IV estimates of adverse child labor effects remain smaller than OLS with no controls, implying negative selection into work based on ability.

VIII. Heterogeneity in Effects of Child Work – Pooled Results

A weakness of the existing literature on child labor is that effects are studied for specific contexts, driven by limited data availability. But child work means different things in different contexts, and it may have different effects at different ages (Attanasio et al. 2017; Del Boca at al. 2017). Young Lives samples children from four LMIC countries with very different economic contexts, and at ages spanning childhood to late adolescence. This gives us an opportunity to go beyond impacts in a particular sample to say something about how the effects of child work vary by context. Here we examine how effects of child work differ across four dimensions: age of the child, level of hours worked, gender and country. As we will see below, we find evidence of heterogenous effects in all four dimensions.

In Section VI we ran separate regressions for children in each age/country cell to avoid any bias due to inappropriate pooling. But we refrained from commenting on the age/country specific coefficients, due to the risk of Type I error when running 28 regressions. In this section, we pool the data across either the country or age dimension (or both) in order to gain efficiency, enabling us to draw more reliable inferences about heterogeneous effects. Pooling also lets us study age/country specific effects in a way that is less susceptible to Type I error.

VIII.A. *Pooled Results*

We start by assessing whether our main results are robust to pooling. Table 7 column 1 presents estimates of equation (8) obtained by pooling the data across both the age and country dimensions. This model includes age and country fixed effects, but all slope

coefficients are assumed homogeneous. Study time is the omitted time-use category.

The pooled results are very similar to the disaggregated results from Section VI. The coefficients on chores and market/farm work are -0.061 and -0.060, respectively, both with standard errors of 0.06. Thus, our estimates imply an extra hour of market/farm work that detracts from study time reduces math scores by 0.06 standard deviations.[43] Obviously, the estimated effect of chores is almost identical.

Furthermore, the coefficient on leisure time is nearly identical to those on market/farm work and chores. Thus, our pooled results strongly support our conclusion in Section VI that market/farm work and chores are equally detrimental for child development if they substitute for study time, while neither is detrimental if it only substitutes for leisure.[44]

The coefficient on school time in the pooled model is small (-0.020) but precisely estimated (standard error 0.007). This supports our result in Figure 7 that school is slightly less productive for child development than study.

Reassuringly then, our pooled results for math are very similar to our disaggregate results. It appears that any misspecification due to pooling is not serious enough to alter our main findings.[45] So we now turn to analysis of differences in effects of child work by country, age and gender. Before we proceed, recall that only our math measure is comparable across age-groups *and* countries (the verbal scores are only comparable across age but not language). Thus, the remainder of the analysis is mostly restricted to studying impacts on math skills.

VIII. B. *Differences by Country*

A key aim of the Young Lives study was to investigate child development in *different* LMIC settings. This was the rationale behind selecting four countries from four major regions of the developing world, with diverse socio-economic characteristics. Child work in such different contexts is likely to mean different things. This motivates an investigation of how the effects of child work differ between the study countries.

Our quantitative data do not contain detailed information on the nature of work, but qualitative data that was also collected as part of the Young Lives study reveals differences across countries/contexts.[46] For example, in Ethiopia, India and Vietnam the dominant work

---

[43] The point estimates in the pooled model are very similar to the median across ages and countries of the point estimates from the disaggregated "full" models in Table 4, i.e. -.055 for market/farm work, -.048 for chores.

[44] Our results also imply that study time is more productive for child cognitive ability than leisure.

[45] Pooled fixed effects results (available on request) are also very similar.

[46] Young Lives conducted in-depth qualitative research on a sub-sample of 200 children, selected to reflect main ethnic/caste groups, rural and urban sites, and poor and less-poor communities (see www.younglives.org.uk).

activities outside the household are linked to agricultural tasks, while in Peru it is informal trade (Morrow and Boyden, 2018). Recall that the Peru sample is much more urban.

Table 7 columns 2-5 reports results by country. Consistent with our general strategy in this section, we pool across ages in order to increase efficiency and reduce the number of country-specific parameters that we estimate down to a manageable number. Given that we pool across ages, we do include age fixed effects.

Our results imply that child work is most detrimental (relative to study) in Ethiopia. It is least detrimental in Peru, where the coefficients on both market/farm work and chores are only about 1/3 as large. India and Vietnam lie in between.[47] Ethiopia is the poorest country in the sample, and the children work quite a bit more than in the other three countries, especially at ages 8 and 12 (see Tables 2 and 3). Households are more reliant on child work in Ethiopia than in the other countries, which may translate into children having to do more physically demanding tasks. Child work may also be more detrimental at the margin simply because the typical child works much higher hours in Ethiopia.

However, as we saw in Table 3, children in Peru tend to work more than in Vietnam or India, especially at ages 8 and 12. Thus, there is no simple relationship whereby the negative impact of child work is greater at higher levels of hours. The relatively benign impact of work in Peru is more likely explained by the different nature of work.

In each of the four countries, we cannot reject the hypothesis that effects of market/farm work, chores and leisure (relative to study) are all equal. Thus, our country level results support our main finding that work activities are only detrimental for child development if they detract from school/study time (not just leisure time).

An additional finding is that school time in Ethiopia and India are less productive than study time. But there is no evidence that school is less productive than study in Vietnam or Peru. In Section VI, when discussing Figure 7, we noted there was some evidence that school is less productive than study, and now we see how that result is driven by Ethiopia and India. Our results are consistent with those of Singh (2020), who presents evidence that Ethiopia and India have lower school productivity than Vietnam and Peru (see also Singh 2015).

VIII.C. _Differences by Child Age_

Table 8 columns 1-4 presents results by age. Here we pool across the four countries, and include country fixed effects. Looking at the age pattern of the coefficients, there is no

---

[47] Formally testing for differences across countries, we find effects in Ethiopia are significantly larger than in the other three countries, but the differences between India, Peru and Vietnam are not statistically significant.

clear pattern of the effect of chores differing by age. However, the point estimates imply that market/farm work has more detrimental effects on child development at ages 8 and 12 than at ages 15 and 18. The estimated effects of market/farm work (relative to study) are about 50% greater at the younger ages, and the differences are statistically significant.

The bottom rows of Table 8 present F-tests for the hypothesis that effects of market/farm work, chores and leisure (relative to study) are equal. There is no evidence against this hypothesis at ages 15 or 18. But there is evidence against it at ages 8 and 12, where the detrimental effect of market/farm work is significantly greater than that of leisure.

There are two distinct explanations for why child work may have a more negative effect on child development at younger ages.[48] First, consistent with findings of Cunha and Heckman (2008), investments in child cognitive development may be more productive at younger ages. This means, conversely, that activities that reduce school/study time by any given number of hours will have a larger negative impact if they occur at younger ages, simply because investment oriented time uses are more productive at younger ages.

Second, it may also be the case that a given amount of time at work is more tiring for younger children. In fact, reports in the qualitative data suggest that some of the work that children are typically expected to do can be especially physically demanding and tiring for young children. As an example, in Ethiopia and India young children's responsibilities can include operation of irrigation pumps on the family farm.[49] This has to be done at night when the electricity supply is reliable, and requires speed and strength to control and direct fast-moving water. Children report being tired at school as a result (Morrow et al, 2014). The qualitative data also reveal substantial age differences in the type of work that children do.

VIII.D. *Differences by Level of Work Hours*

It seems intuitive that the detrimental effect of child work may increase nonlinearly with hours. The qualitative data provide evidence that, beyond a certain threshold, work starts to not just directly crowd-out time for other activities (like school and study) but also impinge on the children's ability to be productive while studying and going to school. In particular, children often report being too tired to learn and pay attention at school because of long work hours, and gradually falling behind in their studies as a result (Zharkevich et al, 2016). In this section we test for non-linearities in the effects of child work.

---

[48] Differences in the effect of child work on child development by age are also consistent with evidence that the same input can have different effects on child development at different ages (e.g. Attanasio, 2015).
[49] Recall that we categorize work on the family farm as part of "market/farm work" rather than chores.

Specifically, we estimate models that include square terms of time spent on domestic chores and market/farm work. At age 8 our estimates are very imprecise, because few children work high levels of hours, making non-linearities very difficult to identify. At ages 15 and 19 we do obtain precise estimates, but we find no clear evidence that the detrimental effect of work increases *more than linearly* with the level of hours. However, we do find clear evidence of non-linearity at age 12. This appears consistent with the evidence above that child work is more detrimental for younger children.

Figure 12 reports the marginal effects of hours of work plotted against the level of hours. The top panel shows age 12 results for math scores. The results show that the negative effects of hours of chores and market/farm work both increase with the level of hours spent in these activities. For example, at zero hours of chores, the marginal effect of the first hour of chores (substituting for study) is only -0.025. But at 3 hours of chores the marginal effect of an additional hour is twice as big.

Estimation of non-linearities in the effect of work *at a point in time* does not require tests scores that are consistent across ages/countries, so the bottom panel of Figure 12 reports results for verbal scores. Here we find an interesting contrast: while the marginal effect of chores increases significantly with the level of chores, we find no evidence of non-linearity in the effect of market/farm work. It is possible that market/farm work stimulates language ability (see Patrinos and Psacharopoulos, 1997), mitigating its negative effects. But we leave further investigation of the mechanism behind this result for future research.

VIII.E. *Differences by Gender*

In Table 8, columns 5 and 6, we present separate results by gender. Here we pool across ages and countries, and include age and country fixed effects. While there is no evidence of gender differences in the effects of time spent on domestic chores (vs. study), the results clearly show that market/farm work is relatively more detrimental for boys. The effect size is more than twice as high for boys than girls (-.068 vs. -.031) and the difference is highly statistically significant.[50] We also find that, for boys, time spent on market/farm work has a significantly more negative effect than time spent in leisure.

The finding that market/farm work is more detrimental for boys is notable, as boys are more likely to engage in market/farm work outside the home than girls. In our data, boys spend 1.47 hours per day (on average) on market/farm work compared to 0.84 for girls. But

---

[50] If we look only at children aged 8 to 12, we find the negative effects of market/farm work increase to -.091 for boys (s.e. = .010) and to -.057 for girls (s.e. = .011). Results available on request.

girls spend 2.20 hours per day on chores, compared to only 1.28 hours for boys.

Consistent with the time-use data, the qualitative data suggest that in most study areas girls have more limited opportunities to engage in market/farm work than boys, due to concerns about their security, and a desire to prepare them for their adult roles (i.e., domestic chores). These restrictions mean boys tend to undertake the more physically demanding and risky market/farm work across the four countries. For example, Morrow and Vennam (2012) note that, in paddy operation in India, more physically demanding tasks like ploughing, sowing, weeding, watering/irrigating, and thrashing are more likely to be done by boys. Hence, an extra hour of market/farm work is likely to be more tiring for boys than girls.

However, a key implication of our results is that, as girls spend almost twice as much time on chores as boys, an exclusive focus on market/farm work, based on the incorrect "status quo" result that chores are not detrimental for development, is *de facto* discriminatory against girls. Indeed the point estimates in Table 8 imply chores are more detrimental for girls than market/farm work. Thus, an exclusive focus on reducing market/farm work ignores the potential benefit of policies to shift girls' time from chores to school/study.

IX. Conclusion

Is any use of children's time better than working, making reduced child work a paramount aim? Or are some time reallocations more productive than others, implying policy should target them specifically? The available evidence does not answer these questions, because it fails to control for the complete time budget of children. We have estimated effects of child work on child development in a framework that accounts for the complete time budget of children, including time spent on chores, market/farm work, school, study and leisure. This allows us to estimate effects of work relative to *specific* alternative time uses.

Our analysis is made possible by a unique multi-country longitudinal cohort study, Young Lives, which collects information on children's time-use along with state-of-the-art measures of cognitive skills. The skill measures are collected regardless of whether children are in school, avoiding an important source of selection bias in prior studies. And the panel dimension of the data allows us to estimate value added (VA) production functions in which unobserved ability and omitted inputs are proxied by lagged achievement scores and rich family background measures. We also report fixed effects and instrumental variables estimates as robustness checks on the VA results.

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), there is little

evidence that either market/farm work or domestic chores are detrimental to child cognitive development if they only substitute for leisure time. However, we find clear evidence that both domestic chores and market/farm work are detrimental to cognitive development if they crowd out school time, and even more detrimental if they crowd out study time.

Our approach yields different findings from the "status quo" approach of estimating effects of child work *without* controls for time spent in other activities. Had we done that, we would have found, as in many prior studies, that time spent on market/farm work is harmful for child development, while time spent on chores is much less so. This would be consistent with the current policy focus on reducing children's time in market/farm work *per se*.

Controlling for the full time-budget, however, we find that *both* domestic chores and market/farm work are detrimental for child development if they crowd out school and study time, but relatively 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 different conclusions from our approach vs. the "status quo" approach stem from differences in the activities that domestic chores and market/farm work typically crowd out. On average, time spent on market/farm work is associated with greater reductions in school time than time spent on domestic chores, especially at higher levels of time.

A weakness of the existing literature on child labor is that effects are studied for specific contexts, driven by limited data availability. Child work means different things in different contexts, and it may have different effects at different ages (Attanasio et al. 2017; Del Boca at al. 2017). The Young Lives data let us study how effects of child work differ across different contexts, as defined by country, age, gender and the level of work. This is the first study to examine cross-country, age and gender differences in effects of child work on child development using outcome measures that are comparable across ages/countries.

Our results imply that child work is most detrimental in Ethiopia and least detrimental in Peru. This may be due to the different nature of work, with children in Ethiopia mostly engaged in farm work, while those in Peru are mostly engaged in informal trade.

There is no clear pattern in the effect of chores by age. But market/farm work is more detrimental for younger children (ages 8 and 12) than older children (ages 15 and 18). We also find evidence that the negative effect of child work on child development increases more than linearly with the level of work, but only for the younger children.

We also find that time spent on market/farm work (rather than school/study) is relatively more detrimental for boys, while the reverse is true for girls. Thus, an exclusive

focus on reducing market/farm work (based on a "status quo" result that chores are benign) ignores the potential benefit of policies to shift girls' time from chores to school/study.

Our fixed effects estimates are very similar to our value added results, despite the fact that these estimators give consistent estimates under somewhat different assumptions about the nature of latent ability and unobserved inputs (see Todd and Wolpin, 2003). This gives us added confidence in the robustness of our main conclusions.

We also attempt to obtain IV results using agricultural prices, wages and weather to instrument for child time-use. This is very challenging because we need to instrument for four endogenous time use variables.[51] We succeed for Ethiopia at ages 8 and 12, where our instruments are strong because child labor is very prevalent and mostly agricultural.

The IV results are consistent with VA in that we find market/farm work and chores have similar negative effects relative to school/study. But the IV results differ from VA in that they imply *larger* negative effects of both market/farm work and chores – both relative to school/study and, for math, relative to leisure. Under a LATE interpretation, this may be because work/chores are particularly detrimental for "marginal" children pushed in and out of child labor by the variation of the instruments. Nevertheless, the IV estimates of adverse child labor effects remain smaller than what we obtain by running OLS on the time-use vector with no controls, implying negative selection into work based on ability.

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 labor without also increasing educational opportunities are unlikely to enhance human capital accumulation.

Our results on the importance of school/study time complement a growing literature that uses time use data and panel data methods to assess the impact of children's time use on child development in developed country contexts (see Cano at al 2019, Del Boca et al 2017, Del Bono et al 2016, Fiorini and Keane, 2014, Funk and Kemper 2016, Hsin and Felfe, 2014, Hull, 2017). This literature generally finds that education-intensive uses of child time are most productive for the development of cognitive skills. Our findings extend this result to a range of low to middle income country contexts.

References

---

[51] We are aware of only three prior studies that attempt to infer the effect of child work on cognitive test scores using IV; namely, Gunnarsson et al. (2009), Bezerra et al (2009) and Dumas (2012). But these studies do not attempt to instrument for the complete vector of child time use.

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, O., E. Fitzsimmons, A. Gomez, M.I. Gutierrez, C. Meghir, and A. Mesnard. 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, O. 2015. "The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies." *Journal of the European Economic Association* 13(6): 949-997.

Attanasio, O., C. Meghir, E. Nix, and F. Salvati. 2017. "Human Capital Growth and Poverty: Evidence from Ethiopia and Peru." *Review of Economic Dynamics* 25:234–59.

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.

Beegle, K. R. Dehejia and R. Gatti. 2006. "Child Labor and Agricultural Shocks," *Journal of Development Economics*, 81: 80-96.

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. *IZA Discussion Paper 4062*. Available at://ftp.iza.org/dp4062.pdf.

Boozer, M. and T. Suri. 2001. "Child Labor and Schooling Decisions in Ghana," Working paper, Yale University.

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.

Cano, Tomás, F. Perales and J. Baxter. 2019. "A Matter of Time: Father Involvement and Child Cognitive Outcomes," *Journal of Marriage and the Family*, 81(1):164-184.

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

Cunha, Flavio and J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Journal of Human Resources*, 43:4, 738-782.

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.

Del Boca, Daniela, Chiara Monfardini and Cheti Nicoletti. 2017. "Parental and Child Time Investments and the Cognitive Development of Adolescents," *Journal of Labor Economics*, 35(2):565-608.

Del Bono, Emilia, Marco Francesconi, Yvonne Kelly and Amanda Sacker. 2016. "Early Maternal Time Investment and Early Child Outcomes," *Economic Journal*, 126:F96-F135.

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." Circle Pines (Minnesota): AGS, American Guidance Service.

Edmonds, Eric V. 2007. "Child Labor." *Handbook of development economics* Vol. 4. Elsevier. pp.3607-3709.

Edmonds, Eric V. 2009. "Defining Child Labour : A Review of the Definitions of Child Labour in Policy Research." Working Paper. International Labour Office, International Programme on the Elimination of Child Labour (IPEC) – Geneva.

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

Fiorini, M. and M. Keane. 2014. "How the Allocation of Children's Time Affects Cognitive and Noncognitive Development", *Journal of Labor Economics* 32(4):787-836.

Funk, Peter and Thorsten Kemper. 2016. "Leisure and Learning Activities and Their Effects on Child Skill Development," University of Cologne, Economics Working Paper #85.

Glewwe, P. and K. Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." *Handbook of the Economics of Education,* Eds E. A. Hanushek, S. Machin, and E. Woessmann Vol. 5: pp. 653–743, Elsevier.

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

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.

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

Hsin, Amy and Christina Felfe. 2014. "When Does Time Matter? Maternal Employment, Children's Time with Parents, and Child Development," *Demography*, 51:1867-94.

Hull, Marie. 2017. "The Time-varying Role of the Family in Student Time Use and Achievement," *IZA Journal of Labor Economics*, 6(10):1-20.

International Labour Organization. 2017a. Global Estimates of Child Labour: Results and Trends, 2012-2016.

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

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.

Keane, M. and K. Wolpin. 1997. "The Career Decisions of Young Men". *Journal of Political Economy* 105(3):473-522

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.

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.

Morrow, V. and U. Vennam. 2012. "Children's Responses to Risk in Agricultural Work in Andhra Pradesh, India." *Development in Practice* 22(4): 549–561

Morrow, V. and J. Boyden. 2018. "Responding to Chidlren's Work: Evidence from the Young Lives study in Ethiopia, India, Peru and Vietnam". Summative Report. Oxford: Young Lives.

Morrow, V., Y. Tafere and U.Vennam . 2014. "Changes to Rural Children's Use of Time: Evidence from Ethiopia and Andhra Pradesh, India" In Bourdillon M & Boyden J (eds) (2014) *Growing up in poverty: findings from Young Lives*. Palgrave, Basingstoke.

Muralidharan, K. and S. Venkatesh. 2013. "Contract Teachers: Experimental Evidence from India." NBER Working Paper No. 19440.

Orazem, Peter and L. Victoria Gunnarsson. 2004. "Child Labour, School Attendance and Performance: A Review." Iowa State University Department of Economics Working Paper # 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.

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.

Sanderson, E. and F. Windmeijer. 2016. "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables," *Journal of Econometrics*, 190, 212-221.

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. 2015. "Private School Effects in Urban and Rural India: Panel Estimates at Primary and Secondary School Ages," *Journal of Development Economics*, 113:16-32.

Singh, Abhijeet. 2020. "Learning More With Every Year: School Year Productivity and International Learning Divergence," *Journal of the European Economic Association*, 18:4, 1770-1813.

Todd, P. and K. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal*. Vol. 113 Issue 485, pp. F3-F33.

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.

Trinh, T. A. Posso and S. Feeny. 2020. Child Labor and Rainfall Deviation: Panel Data Evidence from Rural Vietnam, *The Developing Economies*, 58:1, 53-76.
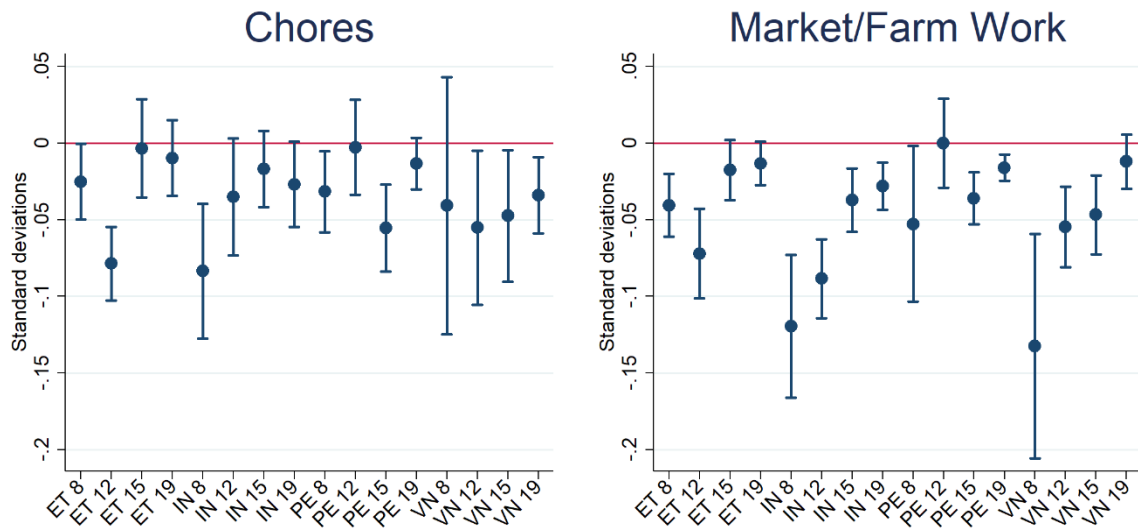
Zharkevich, I, J. Roest and V.T.T.Huong. 2016. "Gendered Trajectories through School, Work and Marriage in Vietnam" *Young Lives Working Paper 156*.

Figure 5: Effect of Child Work on Math Scores

Panel A – Leisure Omitted
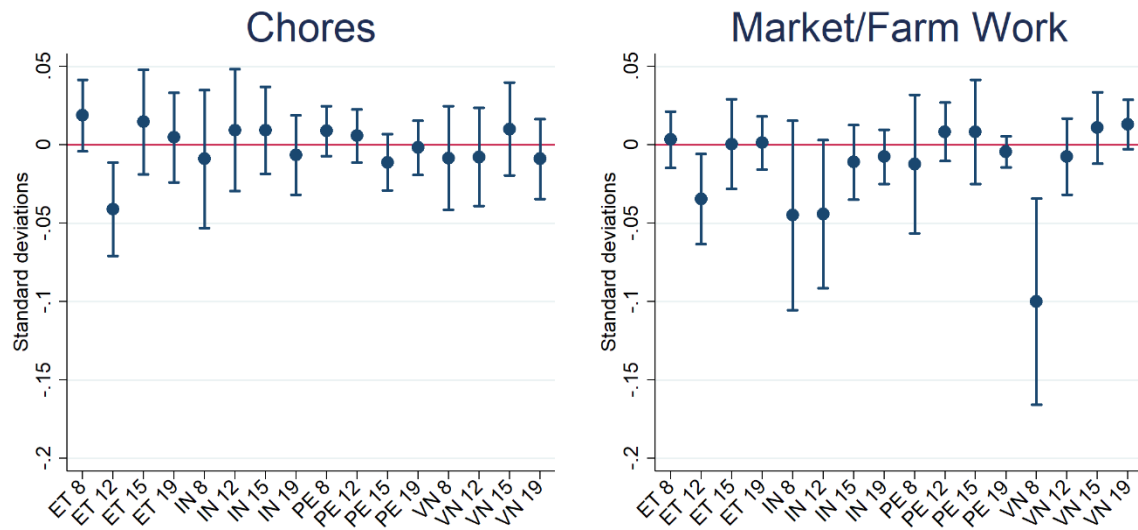
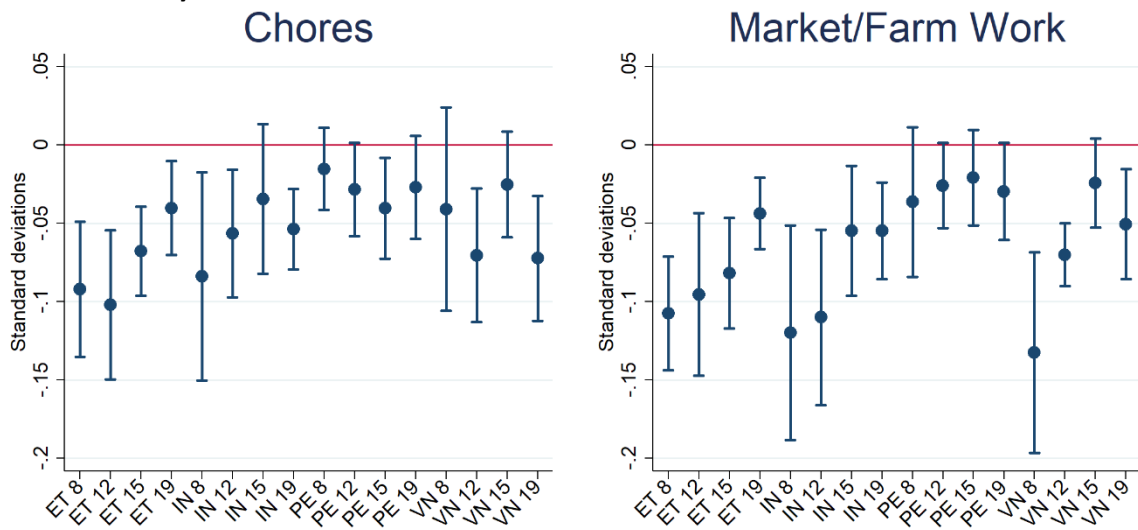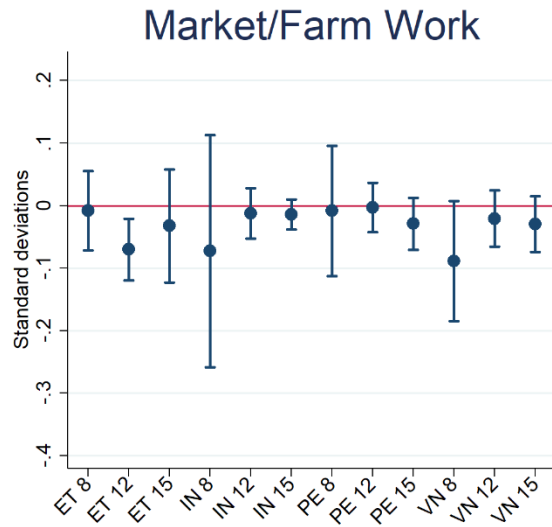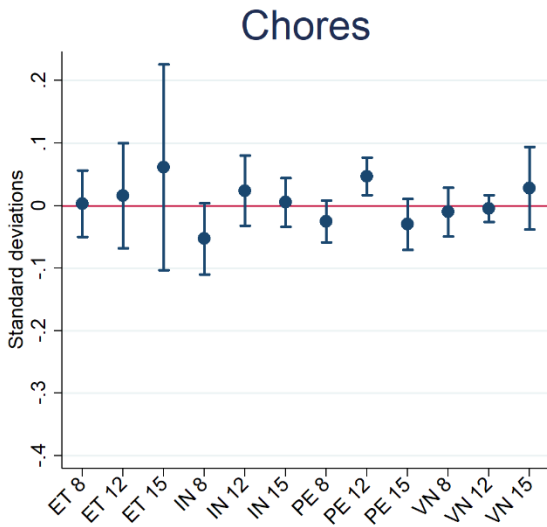**Panel B – School Omitted**
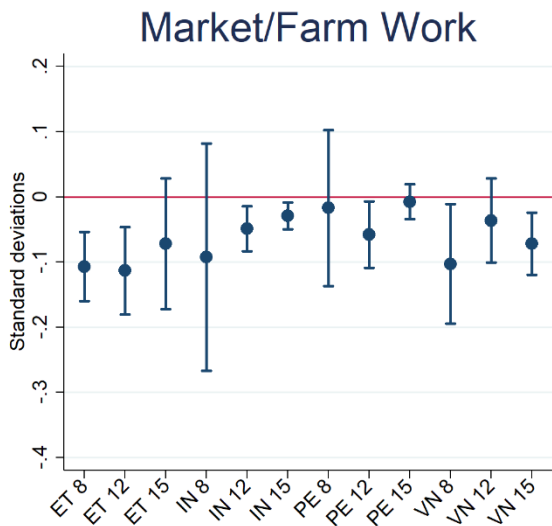

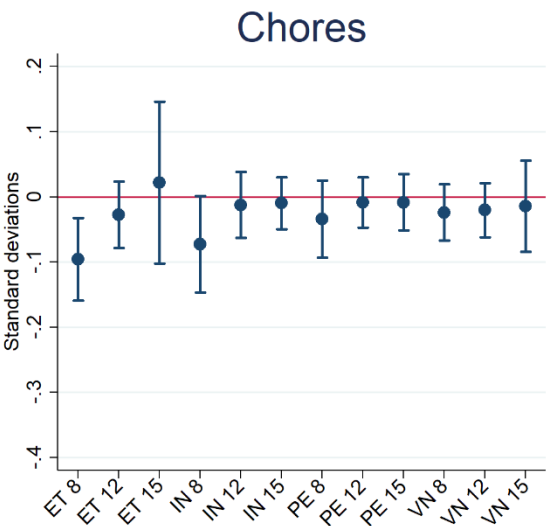**Panel C – Study Omitted**

Figure 6: Effect of Child Work on Verbal Scores

Panel A – Leisure Omitted

Panel B – School Omitted

Panel C – Study Omitted
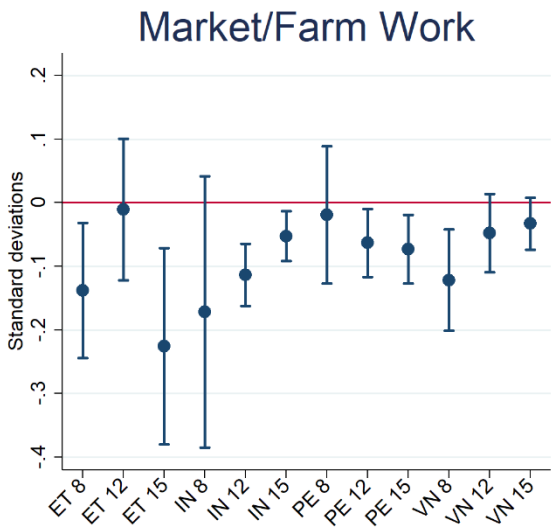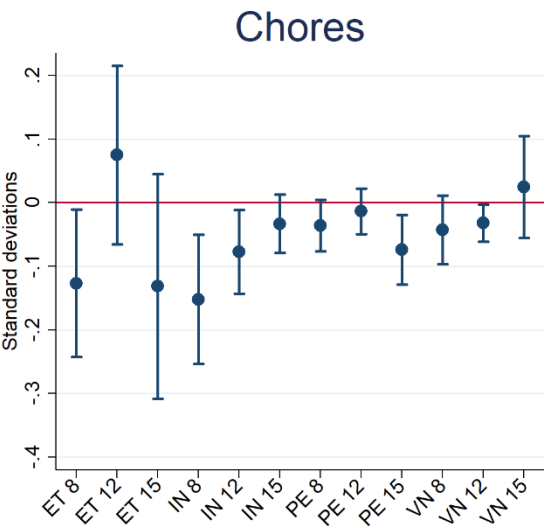
Figure 7: Effect of Schooling on Math and Verbal scores (relative to Study)
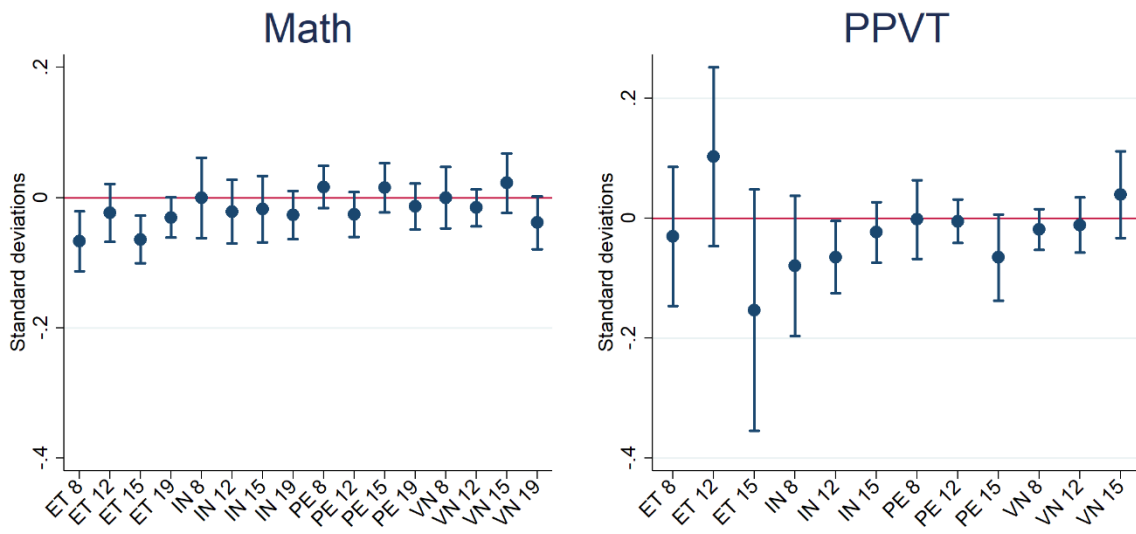
Figure 8: Effect of Market/farm work on Math and Verbal scores (relative to Chores)
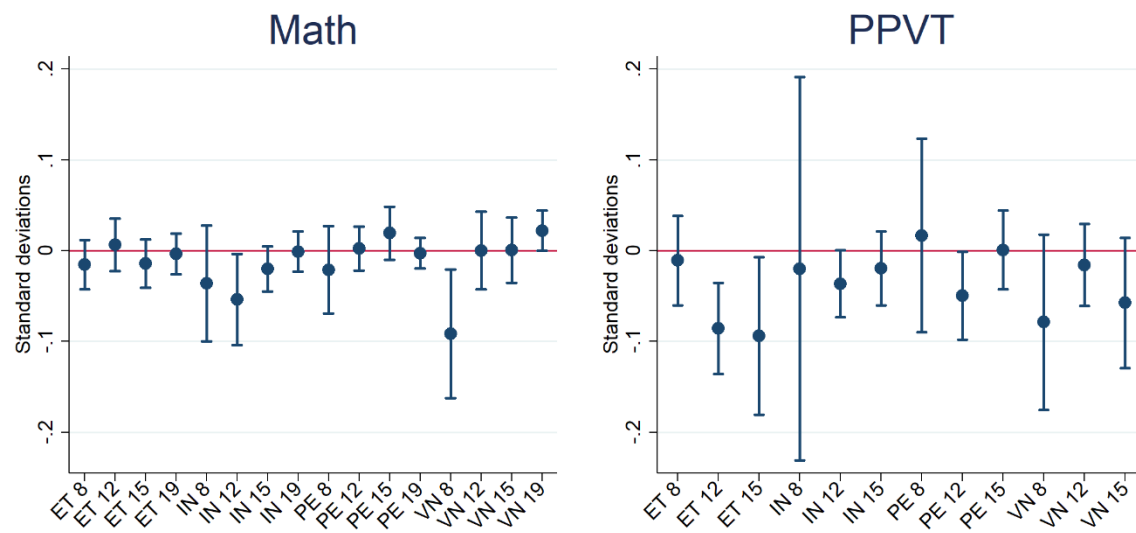
Figure 9: "Status Quo" Analysis of Impact of Work Hours on Math and Verbal scores

Panel A – Math Skills
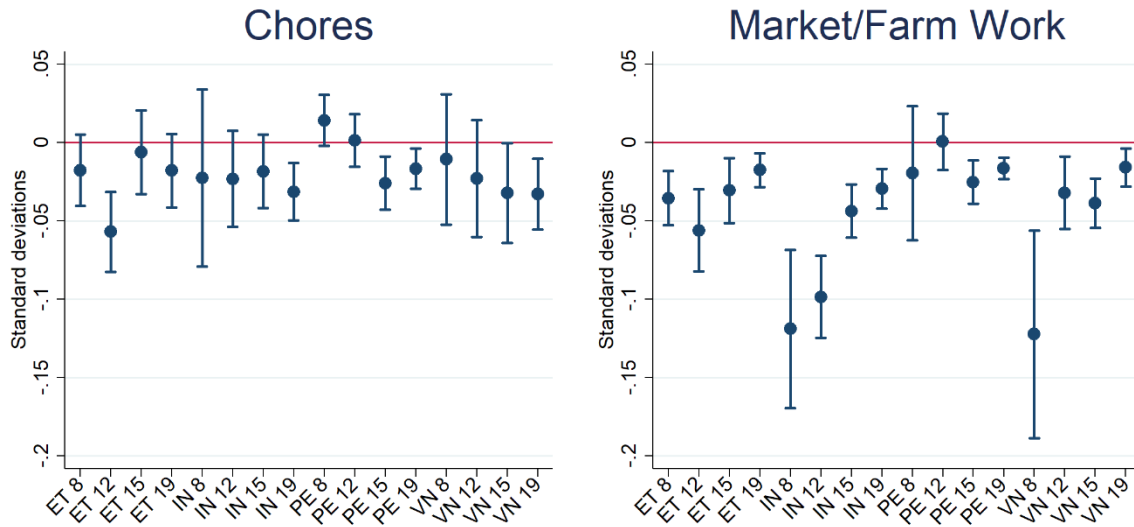
Panel B – Verbal Skills
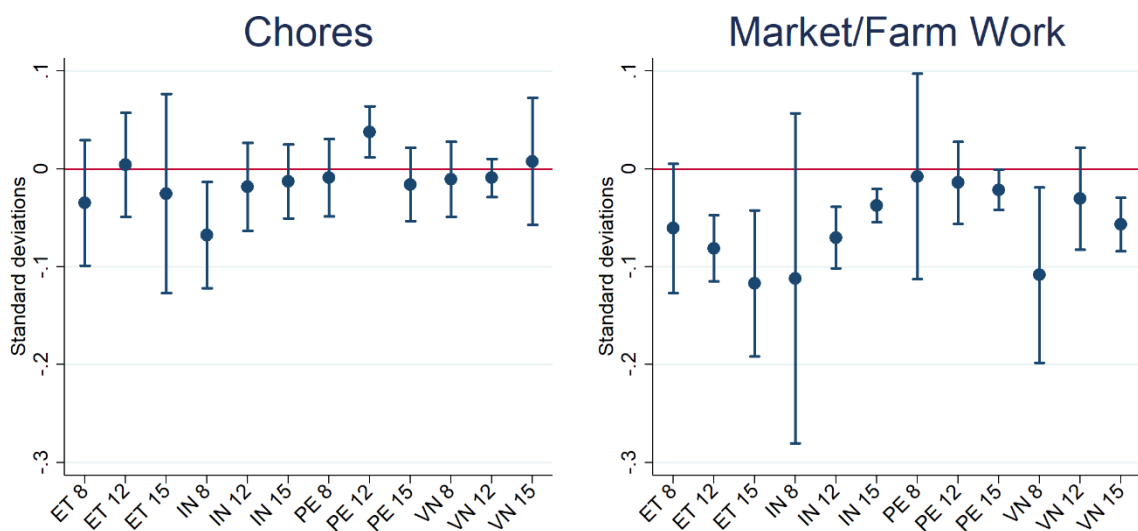
Note: Coefficients are from regressions that control only for hours of work in market/farm work and in household chores. This "status quo" approach contrasts with our main models that control for time allocated to six possible time-use categories that make up a complete 24 hours.

Figure 10: "Status Quo" Analysis of Impact of Child Work on Math and Verbal scores

Panel A – Math Skills
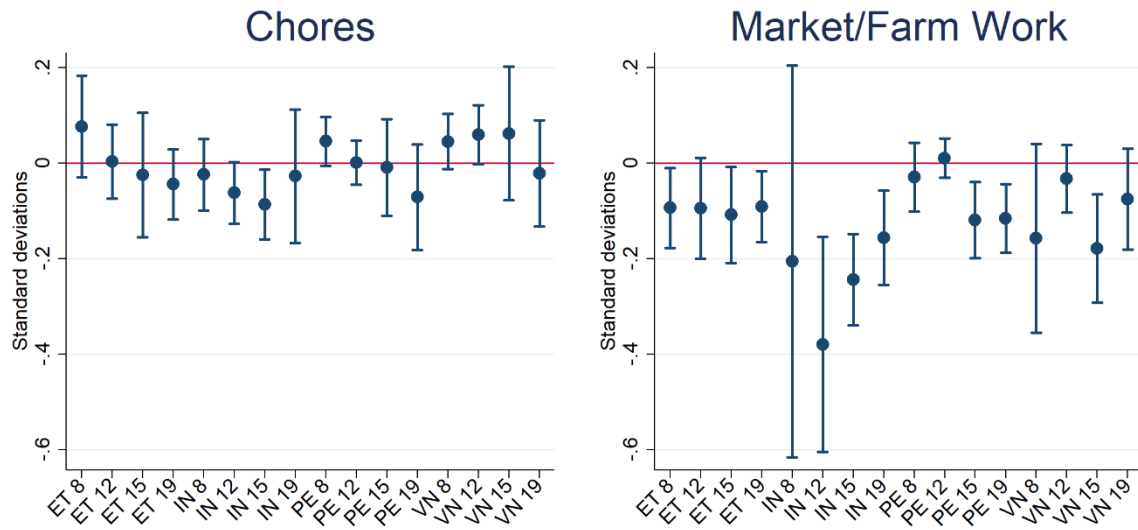
Panel B – Verbal Skills
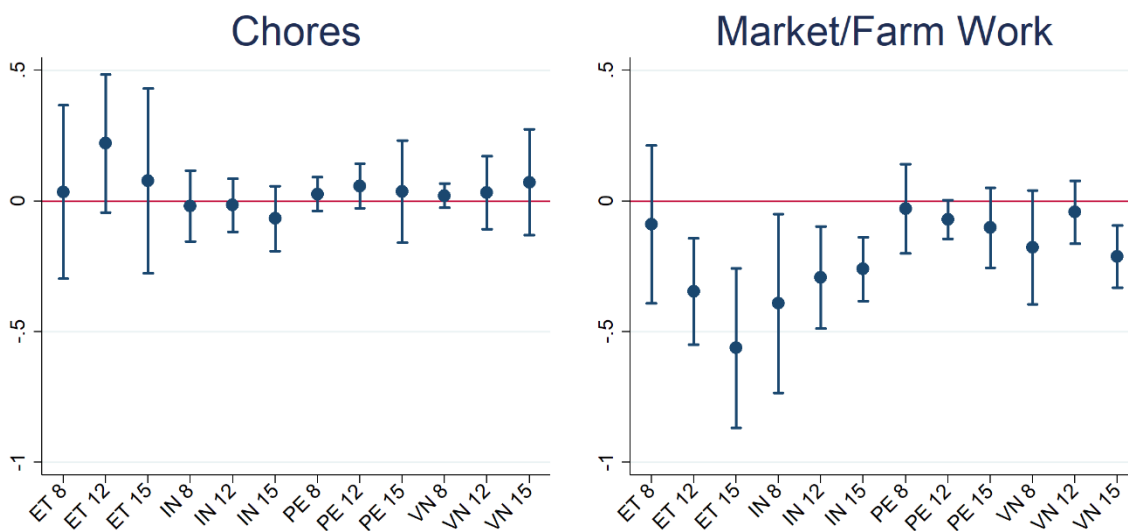
Note: Coefficients are from regressions that control only for indicators of whether the child engages in any market/farm work/household chores. This "status quo" approach contrasts with our main models that control for time allocated to six possible time-use categories that make up a complete 24 hours.

Figure 11: Fixed Effects Results for Math Scores

Leisure omitted

Chores
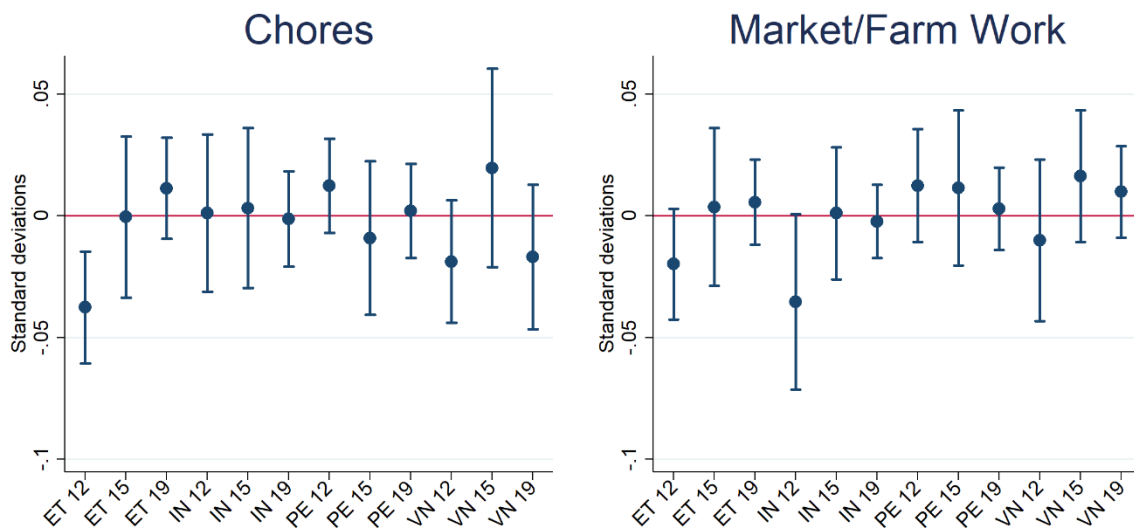
Market/Farm Work

School/Study Time omitted

Chores

Market/Farm Work

Figure 12: Nonlinear Effects of Time on Chores and Market/farm work at Age 12

Panel A – Math Skills

Panel B – Verbal Skills

Note: The figure shows marginal effects of an additional hour of chores or market/farm work at different levels of hours. We present the results for Age 12 only, as we did not find evidence of nonlinearity at later ages.

Table 4: Impact of Control Variables for Math Scores

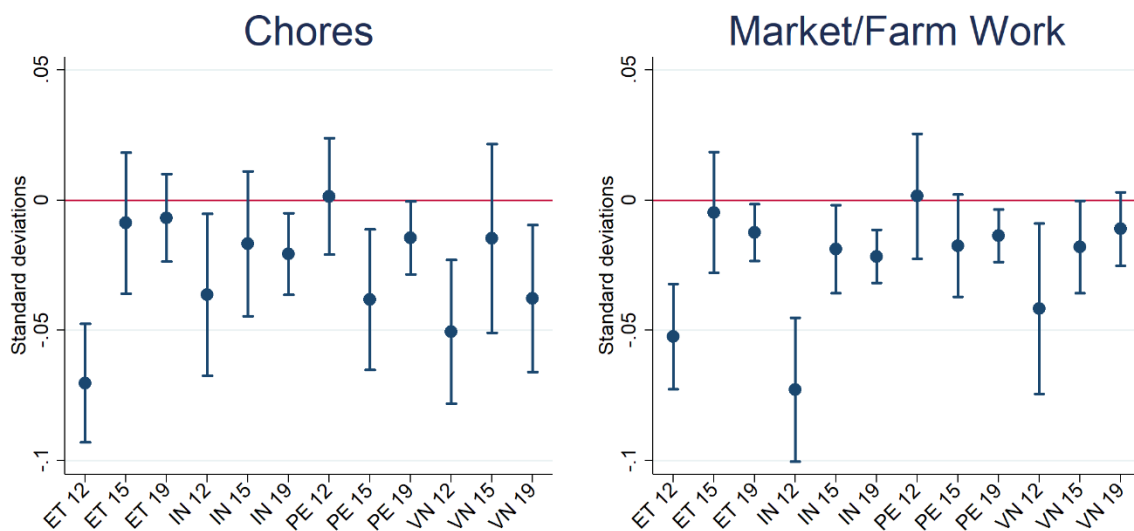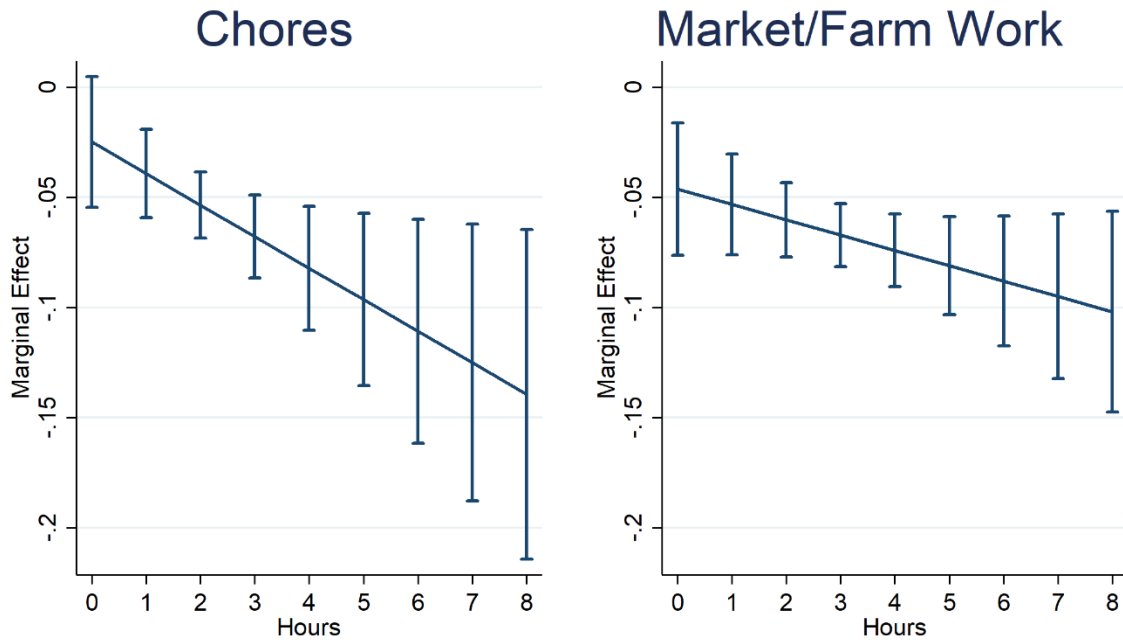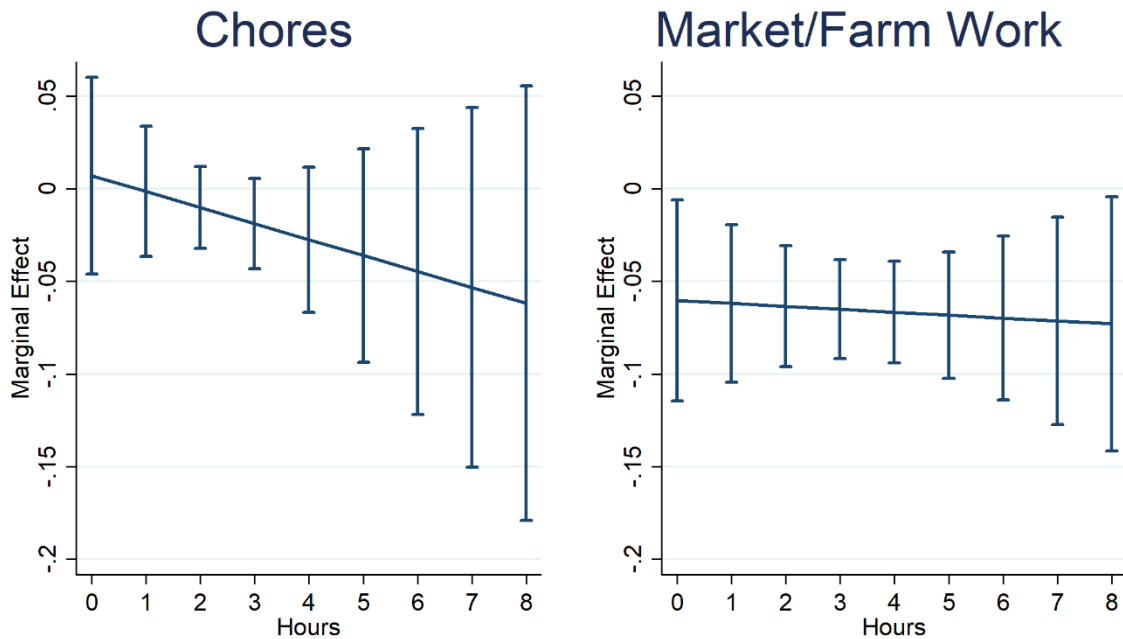|	Chores			Market/Farm work		
	No Controls	Background Controls	VA	No Controls	Background Controls	VA
Age 8						
Ethiopia	-0.168***	-0.097***	-0.092***	-0.182***	-0.109***	-0.108***
	(0.034)	(0.021)	(0.021)	(0.032)	(0.017)	(0.017)
India	-0.179***	-0.093**	-0.084**	-0.138**	-0.123***	-0.120***
	(0.037)	(0.034)	(0.032)	(0.051)	(0.036)	(0.033)
Peru	-0.144***	-0.023	-0.015	-0.199***	-0.047*	-0.036
	(0.017)	(0.015)	(0.012)	(0.029)	(0.023)	(0.023)
Vietnam	-0.168***	-0.041	-0.041	-0.315***	-0.142***	-0.133***
	(0.048)	(0.032)	(0.031)	(0.043)	(0.033)	(0.031)
Age 12						
Ethiopia	-0.289***	-0.161***	-0.102***	-0.262***	-0.150***	-0.095***
	(0.041)	(0.025)	(0.023)	(0.036)	(0.026)	(0.025)
India	-0.101**	-0.080***	-0.057***	-0.154***	-0.152***	-0.110***
	(0.036)	(0.026)	(0.020)	(0.030)	(0.028)	(0.027)
Peru	-0.113***	-0.040**	-0.028*	-0.131***	-0.036**	-0.026*
	(0.021)	(0.015)	(0.014)	(0.024)	(0.014)	(0.013)
Vietnam	-0.181***	-0.083***	-0.070***	-0.227***	-0.098***	-0.070***
	(0.028)	(0.020)	(0.020)	(0.035)	(0.012)	(0.010)
Age 15						
Ethiopia	-0.187***	-0.098***	-0.068***	-0.160***	-0.109***	-0.082***
	(0.022)	(0.017)	(0.014)	(0.026)	(0.018)	(0.017)
India	-0.131***	-0.067**	-0.035	-0.130***	-0.100***	-0.055**
	(0.031)	(0.026)	(0.023)	(0.031)	(0.023)	(0.020)
Peru	-0.121***	-0.072***	-0.040**	-0.103***	-0.058***	-0.021
	(0.024)	(0.018)	(0.015)	(0.024)	(0.015)	(0.015)
Vietnam	-0.110***	-0.041**	-0.025	-0.118***	-0.040**	-0.024*
	(0.026)	(0.018)	(0.016)	(0.027)	(0.017)	(0.014)
Age 19						
Ethiopia	-0.146***	-0.101***	-0.040**	-0.114***	-0.103***	-0.044***
	(0.020)	(0.016)	(0.014)	(0.015)	(0.014)	(0.011)
India	-0.137***	-0.102***	-0.054***	-0.135***	-0.117***	-0.055***
	(0.016)	(0.016)	(0.012)	(0.020)	(0.017)	(0.015)
Peru	-0.045***	-0.046***	-0.027	-0.049***	-0.049***	-0.030*
	(0.013)	(0.014)	(0.016)	(0.017)	(0.016)	(0.015)
Vietnam	-0.142***	-0.101***	-0.072***	-0.122***	-0.090***	-0.050***
	(0.029)	(0.023)	(0.019)	(0.020)	(0.017)	(0.017)

Note: Omitted time category is study in all regressions. Standard errors are in parentheses. A *** indicates significant at the 1% level. ** indicates 5%. * indicates 10%.

Table 5: Instrumental Variable Results for Math Scores

	Linear Regression			IV (Lagged Test Excluded)			IV (Lagged Test Included)		
	OLS (No controls)	OLS	VA	2SLS	2SLS (Lasso)	GMM (Lasso)	2SLS	2SLS (Lasso)	GMM (Lasso)
Ethiopia Age 8									
Market/Farm Wrk									
β	-0.135	-0.058	-0.055	-0.084	-0.088	-0.078	-0.084	-0.101	-0.091
Std. Err.	0.006	0.007	0.008	0.011	0.014	0.014	0.011	0.015	0.015
F-Statistic				6.722	12.577		6.837	13.395	
Chores									
β	-0.125	-0.053	-0.038	-0.090	-0.114	-0.106	-0.089	-0.121	-0.116
Std. Err.	0.007	0.007	0.011	0.015	0.020	0.020	0.015	0.022	0.022
F-Statistic				4.536	9.868		4.330	10.214	
Leisure									
β	-0.089	-0.056	-0.057	-0.050	-0.047	-0.038	-0.045	-0.037	-0.036
Std. Err.	0.006	0.005	0.007	0.010	0.012	0.012	0.010	0.012	0.012
F-Statistic				4.084	8.623		3.922	11.029	
Instruments				123	24		123	20	
Ethiopia Age 12									
Market/Farm Wrk									
β	-0.196	-0.115	-0.078	-0.136	-0.143	-0.138	-0.092	-0.111	-0.111
Std. Err.	0.009	0.011	0.015	0.017	0.021	0.021	0.017	0.020	0.020
F-Statistic				6.507	15.740		6.099	16.343	
Chores									
β	-0.226	-0.121	-0.084	-0.195	-0.187	-0.149	-0.134	-0.128	-0.097
Std. Err.	0.012	0.012	0.012	0.026	0.030	0.029	0.024	0.029	0.027
F-Statistic				5.124	14.607		4.804	14.579	
Leisure									
β	-0.112	-0.080	-0.044	-0.059	-0.009	-0.014	-0.032	0.009	0.009
Std. Err.	0.011	0.010	0.011	0.020	0.029	0.028	0.018	0.025	0.025
F-Statistic				3.941	6.990		4.431	8.487	
Instruments				132	27		132	24	

Note: Omitted time category is school and study in all regressions. The first stage F-statistic reported is calculated according to Sanderson and Windmeijer (2016). The optimal regularization parameter in Lasso was picked according to the minimal BIC.

Table 6: Instrumental Variable Results for Verbal Scores

	Linear Regression			IV (Lagged Test Excluded)			IV (Lagged Test Included)		
	OLS (No controls)	OLS	VA	2SLS	2SLS (Lasso)	GMM (Lasso)	2SLS	2SLS (Lasso)	GMM (Lasso)
Ethiopia Age 8									
Market/Farm Wrk									
β	-0.271	-0.131	-0.118	-0.184	-0.218	-0.237	-0.170	-0.223	-0.245
Std. Err.	0.012	0.014	0.011	0.024	0.032	0.032	0.024	0.033	0.032
F-Statistic				11.526	15.088		11.890	16.392	
Chores									
β	-0.253	-0.082	-0.080	-0.176	-0.185	-0.194	-0.128	-0.158	-0.179
Std. Err.	0.014	0.016	0.021	0.037	0.058	0.057	0.035	0.063	0.061
F-Statistic				4.011	6.626		4.008	6.678	
Leisure									
β	-0.189	-0.106	-0.089	-0.163	-0.207	-0.209	-0.152	-0.207	-0.210
Std. Err.	0.012	0.011	0.013	0.022	0.031	0.030	0.024	0.038	0.037
F-Statistic				4.054	6.572		3.999	6.621	
Instruments				123	24		123	20	
Ethiopia Age 12									
Market/Farm Wrk									
β	-0.297	-0.150	-0.110	-0.181	-0.172	-0.174	-0.147	-0.138	-0.121
Std. Err.	0.014	0.015	0.021	0.027	0.040	0.039	0.025	0.035	0.034
F-Statistic				6.267	15.203		8.961	17.179	
Chores									
β	-0.212	-0.051	-0.020	-0.126	-0.192	-0.211	-0.054	-0.195	-0.170
Std. Err.	0.018	0.018	0.022	0.039	0.058	0.057	0.038	0.065	0.064
F-Statistic				5.179	11.895		4.714	7.904	
Leisure									
β	-0.122	-0.041	-0.007	-0.095	-0.190	-0.176	-0.029	-0.103	-0.074
Std. Err.	0.017	0.015	0.025	0.030	0.049	0.048	0.027	0.047	0.046
F-Statistic				4.290	8.005		6.200	7.361	
Instruments				132	27		132	24	

Note: Omitted time category is school and study in all regressions. The first stage F-statistic reported is calculated according to Sanderson and Windmeijer (2016). The optimal regularization parameter in Lasso was picked according to the minimal BIC.

Table 7: Pooled Models for Math Scores (Study Omitted): Geographic variation

	Pooled (1)	Ethiopia (2)	India (3)	Peru (4)	Vietnam (5)
Hours/day in School	-0.020***	-0.056***	-0.030**	-0.003	-0.003
	(0.007)	(0.013)	(0.014)	(0.009)	(0.012)
Hours/day spent on Chores	-0.061***	-0.084***	-0.065***	-0.028***	-0.052***
	(0.006)	(0.008)	(0.012)	(0.007)	(0.011)
Hours/day spent on Market/farm work	-0.060***	-0.084***	-0.072***	-0.032***	-0.043***
	(0.006)	(0.008)	(0.012)	(0.007)	(0.009)
Hours/day spent on Leisure	-0.056***	-0.080***	-0.063***	-0.030***	-0.049***
	(0.006)	(0.009)	(0.013)	(0.006)	(0.008)
Adjusted R2	0.687	0.620	0.604	0.692	0.599
Sample Size	20,351	4,821	5,383	4,908	5,239

F-test for equality of chores and leisure coefficients:

F-statistic	1.71	0.37	0.02	0.19	0.08
p-value	0.195	0.551	0.877	0.671	0.781

F-test for equality of market/farm work and leisure coefficients:

F-statistic	1.51	0.39	2.02	0.23	1.07
p-value	0.223	0.539	0.17	0.637	0.314

note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$; study time is the omitted time-use category in all these regressions. Standard errors are in parentheses.

Table 8: Pooled Models for Math Scores (Study Omitted): Child characteristics

	Age 8 (1)	Age 12 (2)	Age 15 (3)	Age 19 (4)	Female (5)	Male (6)
Hours/day spent in School	-0.005	-0.006	-0.023*	-0.022***	0.018	-0.008
	(0.015)	(0.010)	(0.011)	(0.008)	(0.011)	(0.010)
Hours/day spent on Chores	-0.056***	-0.068***	-0.057***	-0.043***	-0.045***	-0.051***
	(0.016)	(0.009)	(0.009)	(0.007)	(0.009)	(0.009)
Hours/day spent Market/farm work	-0.081***	-0.074***	-0.054***	-0.041***	-0.031***	-0.068***
	(0.016)	(0.009)	(0.009)	(0.007)	(0.009)	(0.008)
Hours/day spent on Leisure	-0.054***	-0.056***	-0.056***	-0.042***	-0.037***	-0.054***
	(0.014)	(0.008)	(0.009)	(0.007)	(0.009)	(0.008)
Adjusted R2	0.578	0.623	0.639	0.648	0.644	0.640
Sample Size	6,460	7,325	3,429	3,137	9,844	10,507

F-test for equality of chores and leisure coefficients:

	Age 8 (1)	Age 12 (2)	Age 15 (3)	Age 19 (4)	Female (5)	Male (6)
F-statistic	0.08	2.53	0.01	0.04	2.41	0.22
p-value	0.774	0.116	0.907	0.84	0.124	0.637

F-test for equality of market/farm work and leisure coefficients:

	Age 8 (1)	Age 12 (2)	Age 15 (3)	Age 19 (4)	Female (5)	Male (6)
F-statistic	8.43	5.30	0.07	0.11	1.64	9.11
p-value	0.005	0.024	0.795	0.737	0.204	0.003

note: *** p<0.01, ** p<0.05, * p<0.1; study time is the omitted time-use category in all these regressions. Standard errors are in parentheses.