Supplement to "Child work and cognitive development: Results from four low to middle income countries"

(*Quantitative Economics*, Vol. 13, No. 2, May 2022, 425–465)

Michael Keane
School of Economics, University of New South Wales

Sonya Krutikova
Institute for Fiscal Studies

Timothy Neal
School of Economics, University of New South Wales

Appendix A: Comparative statics in the theoretical framework

This Appendix derives the comparative static results for the simple theoretical framework of Section 3. We work with the two first-order conditions:

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

and ignore $u_c w = u_l$ as we assume for simplicity that parent work is predetermined. We also focus on the special case where $p = 0$, and $u_{YY} = 0$. Total differentiation of the FOCs gives:

$$u_{ss}\,ds + u_{s\mu}\,d\mu + u_{l_c l_c}\,ds + u_{l_c l_c}\,dh_c = 0, \tag{A.1}$$

$$u_{cc} w_c^2\,dh_c + u_{cc} w_c\,dN + u_{l_c l_c}\,dh_c + u_{l_c l_c}\,ds = 0. \tag{A.2}$$

(1) First, to derive $\frac{ds}{d\mu} > 0$,we set $dN = 0$ and simplify (A.2) to

$$u_{l_c l_c}\,dh_c = -\frac{(u_{l_c l_c})^2}{u_{cc} w_c^2 + u_{l_c l_c}}\,ds.$$

Substituting this expression into (A.1), we have

$$\left(u_{ss} + u_{l_c l_c} - \frac{(u_{l_c l_c})^2}{u_{cc} w_c^2 + u_{l_c l_c}}\right) ds = -u_{s\mu} d_\mu.$$

Simplifying

$$\frac{ds}{d\mu} = -\frac{u_{s\mu}}{u_{ss} + u_{l_c l_c} R_c} > 0 \quad \text{where } R_c = \left(u_{cc} w_c^2 / \left[u_{cc} w_c^2 + u_{l_c l_c}\right]\right).$$

Michael Keane: m.keane@unsw.edu.au
Sonya Krutikova: sonya_k@ifs.org.uk
Timothy Neal: timothy.neal@unsw.edu.au

Available at http://qeconomics.org. https://doi.org/10.3982/QE1745

(2) Second, to derive $\frac{dh_c}{d\mu} < 0$, we simplify (A.2) to

$$u_{l_cl_c}\,ds = -\left(u_{cc}w_c^2 + u_{l_cl_c}\right)dh_c.$$

Substituting this expression into (A.1) we have:

$$u_{ss}\,ds + u_{s\mu}d_\mu - u_{cc}w_c^2\,dh_c = 0.$$

Now, multiply (A.2) by $u_{ss}/u_{l_cl_c}$ and simplify

$$\frac{u_{ss}}{u_{l_cl_c}}\left(u_{cc}w_c^2 + u_{l_cl_c}\right)dh_c + u_{ss}\,ds = 0.$$

Substituting for $u_{ss}\,ds$,

$$\frac{u_{ss}}{u_{l_cl_c}}\left(u_{cc}w_c^2 + u_{l_cl_c}\right)dh_c + u_{cc}w_c^2\,dh_c = u_{s\mu}d_\mu.$$

Simplifying we have

$$\frac{dh_c}{d\mu} = \frac{u_{s\mu}}{u_{ss} + u_{cc}w_c^2 + \frac{u_{ss}}{u_{l_cl_c}}u_{cc}w_c^2} \rightarrow \frac{dh_c}{d\mu} = \frac{u_{s\mu}}{u_{ss} + u_{cc}w_c^2R_l} < 0,$$

where $R_l = ([u_{ss} + u_{l_cl_c}]/u_{l_cl_c})$.

(3) Third, to derive $\frac{ds}{dN} > 0$,we set $d\mu = 0$ and simplify (A.1) to

$$dh_c = -\frac{u_{ss} + u_{l_cl_c}}{u_{l_cl_c}}\,ds.$$

Substituting into (A.2) and simplifying yields

$$u_{cc}w_c\,dN = -u_{l_cl_c}\,ds + \frac{\left(u_{l_cl_c} + u_{cc}w_c^2\right)\left(u_{ss} + u_{l_cl_c}\right)}{u_{l_cl_c}}\,ds$$

$$\rightarrow u_{cc}w_c\,dN = \frac{u_{cc}w_c^2u_{ss} + u_{cc}w_c^2u_{l_cl_c} + u_{l_cl_c}u_{ss}}{u_{l_cl_c}}\,ds$$

$$\rightarrow \frac{ds}{dN} = \frac{u_{cc}w_c}{u_{ss} + u_{cc}w_c^2R_l} > 0.$$

(4) Finally, to see that $\frac{dh_c}{dN} < 0$, simplify (A.1) to

$$dS = -\frac{u_{l_cl_c}}{u_{ss} + u_{l_cl_c}}\,dh_c.$$

Then insert into (A.2),

$$u_{cc}w_c\,dN = \frac{(u_{l_cl_c})^2}{u_{ss} + u_{l_cl_c}}\,dh_c - \left(u_{l_cl_c} + u_{cc}w_c^2\right)dh_c$$

$$\rightarrow \frac{dh_c}{dN} = -\frac{u_{cc}w_c}{u_{cc}w_c^2 + \frac{u_{l_cl_c}}{R_l}} < 0.$$

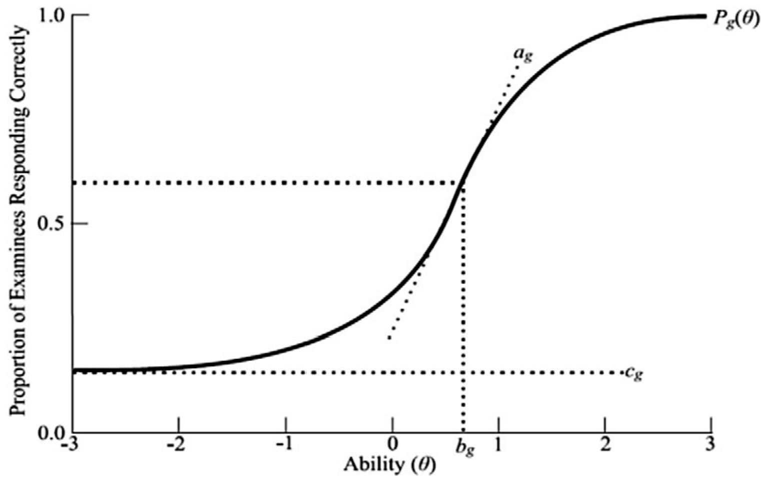


FIGURE B.1. Three parameter item characteristic curve (3PL model).

APPENDIX B: 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, Young Lives uses Item Response Theory (IRT) to model the math and verbal test scores. This allows one to construct estimates of *latent* math and verbal ability that are comparable across ages and (for math) countries. IRT has a long history in education and psychometrics (e.g., van der Linden and Hambleton (1997)). It underlies the construction of standard IQ tests and well-known aptitude tests like the SAT, and 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), Verriest (2020)), 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 B.1, we assume a standard three parameter ICC, where the parameters (a, b, c) map ability into responses.

The parameter c_g in Figure B.1 is the probability a person can guess the correct answer to item g, given he/she has essentially no ability to actually determine the correct answer. Parameter b_g is known as the "difficulty" of item g. It measures the ability level at which the probability of a correct answer lies halfway between the probability of a correct guess and 100%.[1] 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).

[1] 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%.

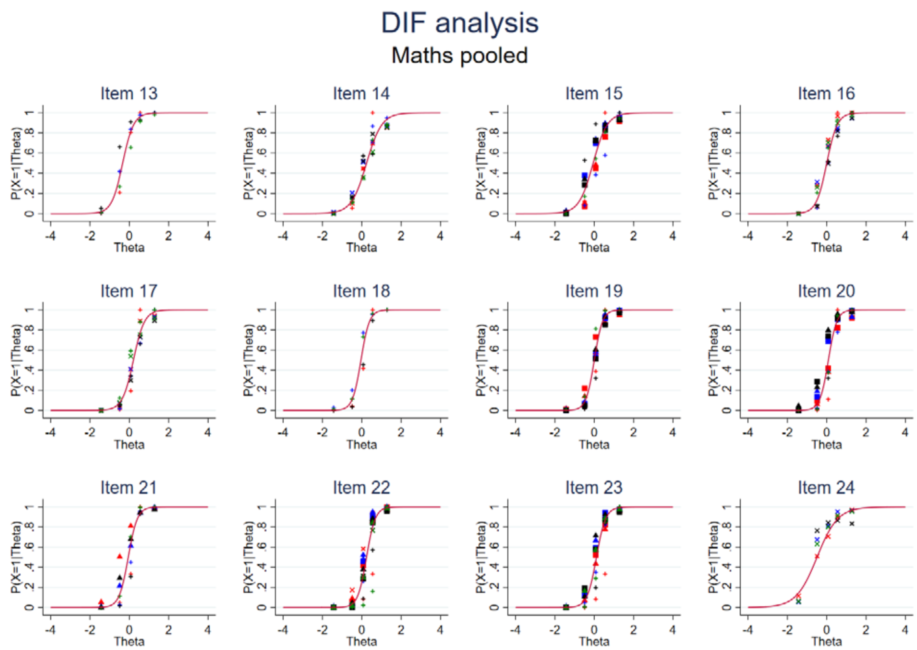


FIGURE B.2. Testing for differential item functioning (DIF). *Notes*: The figure shows ICC curves for twelve items that have overlap across countries/cohorts/rounds. For each context, we also plot the empirical frequency of a correct answer at quintiles of ability. Contexts are coded as: Ethiopia—Red, India—Blue, Peru—Black, Vietnam—Green. Young Cohort: (R3) +, (R4) ▲, Older Cohort: (R3) **x**, (R4) ■. If pooling is acceptable, then all ability points should lie on the ICC for each item. The empirical frequencies corresponding to each quintile are often not clearly visible, as they often lie essentially on top of one another.

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 5–15. 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.

Another key advantage of IRT is that it offers a way of running diagnostics on the cross-cultural comparability of test items. The idea is to check if the ICC curve for an item is invariant to context. “Differential item functioning” (DIF) arises if people from different contexts (based on country/cohort/round) who have the *same* latent ability have *different* probabilities of answering an item correctly. That means the ICC for that item differs by context. In that case pooling questions across contexts is not appropriate.

Figure B.2 presents an example of such a DIF analysis. The complete math question bank contains 104 items that can be linked across a least a subset of the countries, cohorts and ages. The figure reports a DIF analysis for twelve of these items (as space precludes showing all 104). It plots the known item characteristic curves for each item. And it also indicates the rate of correct responses at quintiles of ability in several different country/cohort/round (age) contexts. 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 B.2 shows that the ability quintiles for different countries/cohorts/ages line up along the ICC curves quite accurately in nearly all instances. The results for the verbal scores are similarly encouraging.[2]

Appendix C: Evolution of school hours with market/farm work and domestic chores

Figure 1 in the text shows how the level of school hours varies conditional on hours of market/farm work and hours of domestic chores. The predictions are from a multivariate fractional polynomial model developed in Royston and Altman (1994) and Royston and Sauerbrei (2008), which uses a data-driven model selection algorithm to find the best polynomial fit for multiple variables in a linear regression. In this application,

$$S_{it} = f(W_{it}) + g(C_{it}) + \beta \mathrm{X}_{\mathrm{it}} + e_{it},$$

where S_{it} is hours of schooling, W_{it} is hours of market/farm work, C_{it} is hours of domestic chores, and X_{it} is a vector of controls that are entered linearly. The functions $f(\cdot)$ and $g(\cdot)$ are estimated using an optimized polynomial of W_{it} and C_{it}. Since the final functional form is linear in parameters, it is estimated by OLS.

The optimized model that was used to generate Figure 1 is

$$S_{it} = 0.020 W_{it}^3 \ln(W_{it}) - 0.052 W_{it}^3 + \frac{0.0862}{10{,}000} C_{it}^{-2} - 0.075 C_{it}^2 + \beta \mathrm{X}_{\mathrm{it}},$$

where $N = 19{,}696$. The control variables included here are dummies for each of the 16 different country/round/cohort combinations (to control for age and country level effects), the wealth index, the urban/rural dummy, the number of brothers and sisters, parental education, and gender.

[2]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.

Appendix D: Summary statistics

Table D.1. Summary statistics.

	Ethiopia		India		Peru		Vietnam	
	YC (Age 5)	OC (Age 12)	YC (Age 5)	OC (Age 12)	YC (Age 5)	OC (Age 12)	YC (Age 5)	OC (Age 12)
Male	0.528	0.510	0.534	0.495	0.503	0.533	0.512	0.496
	(0.499)	(0.500)	(0.499)	(0.500)	(0.500)	(0.499)	(0.500)	(0.500)
Mother's Education:								
None	0.496	0.443	0.512	0.582	0.0829	0.101	0.116	0.0954
	(0.500)	(0.497)	(0.500)	(0.493)	(0.276)	(0.301)	(0.320)	(0.294)
primary	0.397	0.484	0.279	0.289	0.454	0.506	0.484	0.493
	(0.489)	(0.500)	(0.449)	(0.454)	(0.498)	(0.500)	(0.500)	(0.500)
secondary	0.0869	0.0617	0.177	0.103	0.279	0.247	0.327	0.363
	(0.282)	(0.241)	(0.382)	(0.304)	(0.448)	(0.432)	(0.469)	(0.481)
post secondary	0.0197	0.0117	0.0321	0.0259	0.184	0.146	0.0730	0.0488
	(0.139)	(0.108)	(0.176)	(0.159)	(0.388)	(0.353)	(0.260)	(0.216)
Father's Education:								
None	0.209	0.171	0.332	0.397	0.0115	0.0136	0.0761	0.0679
	(0.407)	(0.377)	(0.471)	(0.490)	(0.107)	(0.116)	(0.265)	(0.252)
primary	0.628	0.715	0.320	0.333	0.346	0.563	0.460	0.446
	(0.483)	(0.452)	(0.467)	(0.472)	(0.476)	(0.496)	(0.499)	(0.497)
secondary	0.114	0.0851	0.255	0.201	0.452	0.274	0.378	0.406
	(0.318)	(0.279)	(0.436)	(0.401)	(0.498)	(0.446)	(0.485)	(0.491)
post secondary	0.0487	0.0287	0.0925	0.0685	0.190	0.149	0.0855	0.0795
	(0.215)	(0.167)	(0.290)	(0.253)	(0.393)	(0.356)	(0.280)	(0.271)
Household size	6.038	6.532	5.520	5.182	5.490	5.560	4.673	4.900
	(2.058)	(2.042)	(2.232)	(1.826)	(2.072)	(1.970)	(1.514)	(1.391)
No. brothers in house	1.454	1.886	0.652	0.968	0.953	1.286	0.536	0.887
	(1.380)	(1.449)	(0.724)	(0.842)	(1.118)	(1.161)	(0.707)	(0.846)
No. sisters in house	1.384	1.861	0.809	0.890	0.951	1.127	0.616	0.881
	(1.333)	(1.411)	(0.875)	(0.901)	(1.153)	(1.067)	(0.842)	(0.972)
No. grandparents in house	0.238	0.181	0.960	0.507	0.617	0.346	0.680	0.278
	(0.548)	(0.468)	(0.963)	(0.752)	(0.891)	(0.674)	(0.907)	(0.564)
No. other adults in house	0.419	0.406	0.764	0.341	0.709	0.475	0.452	0.148
	(1.067)	(1.013)	(1.424)	(1.054)	(1.271)	(1.044)	(1.003)	(0.542)
No. other elderly in house	0.00547	0.0106	0.00105	0	0.0132	0.0121	0.00573	0.00212
	(0.0738)	(0.103)	(0.0324)	(0)	(0.127)	(0.109)	(0.0755)	(0.0460)
Both parents in house	0.759	0.651	0.945	0.863	0.791	0.709	0.919	0.917
	(0.428)	(0.477)	(0.228)	(0.344)	(0.407)	(0.454)	(0.273)	(0.276)
Mother's age	31.45	38.26	27.63	34.70	31.27	38.40	31.18	38.43
	(6.404)	(6.945)	(4.315)	(5.634)	(6.566)	(6.555)	(5.782)	(5.713)
Father's age	40.72	47.96	33.44	40.89	35.31	42.47	34.09	40.77
	(8.650)	(8.385)	(5.204)	(6.160)	(6.889)	(6.942)	(5.948)	(6.025)
Child's age (in months)	62.37	145.2	64.74	148.5	63.98	148.2	63.67	147.6
	(3.796)	(3.717)	(3.714)	(4.080)	(4.681)	(5.227)	(3.619)	(3.782)
Child lives in urban area	0.401	0.407	0.252	0.245	0.695	0.747	0.205	0.196
	(0.490)	(0.492)	(0.434)	(0.431)	(0.461)	(0.435)	(0.404)	(0.397)

(*Continues*)

Table D.1. *Continued.*

	Ethiopia		India		Peru		Vietnam	
	YC (Age 5)	OC (Age 12)	YC (Age 5)	OC (Age 12)	YC (Age 5)	OC (Age 12)	YC (Age 5)	OC (Age 12)
Height for age z-score	−1.436	−1.354	−1.632	−1.504	−1.466	−1.485	−1.325	−1.464
	(1.098)	(1.209)	(1.013)	(1.137)	(1.182)	(1.108)	(1.073)	(1.063)
Wealth index	0.287	0.303	0.459	0.469	0.470	0.504	0.490	0.512
	(0.178)	(0.169)	(0.195)	(0.200)	(0.230)	(0.222)	(0.181)	(0.171)
Child religion (1)	0.717	0.727	0.875	0.874	0.809	0.837	0.857	0.839
	(0.451)	(0.446)	(0.330)	(0.331)	(0.393)	(0.369)	(0.350)	(0.368)
Child religion (2)	0.114	0.112	0.0699	0.0674	0.133	0.133	0.143	0.161
	(0.318)	(0.315)	(0.255)	(0.251)	(0.340)	(0.339)	(0.350)	(0.368)
Child religion (3)	0.160	0.154	0.0547	0.0581	0.0577	0.0301	NA	NA
	(0.366)	(0.361)	(0.227)	(0.234)	(0.233)	(0.171)		
Child religion (4)	0.00984	0.00745	NA	NA	NA	NA	NA	NA
	(0.0987)	(0.0860)						
Child ethnicity (1)	0.288	0.287	0.181	0.206	0.916	0.926	0.857	0.874
	(0.453)	(0.453)	(0.385)	(0.405)	(0.277)	(0.262)	(0.350)	(0.332)
Child ethnicity (2)	0.214	0.207	0.147	0.109	0.0561	0.0422	0.143	0.126
	(0.410)	(0.406)	(0.354)	(0.312)	(0.230)	(0.201)	(0.350)	(0.332)
Child ethnicity (3)	0.228	0.229	0.469	0.469	0.0278	0.0316	NA	NA
	(0.420)	(0.420)	(0.499)	(0.499)	(0.164)	(0.175)		
Child ethnicity (4)	0.270	0.277	0.139	0.152	NA	NA	NA	NA
	(0.444)	(0.448)	(0.346)	(0.360)				
Child ethnicity (5)	NA	NA	0.0636	0.0633	NA	NA	NA	NA
			(0.244)	(0.244)				
N	1820	939	1896	962	1889	660	1901	933

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

Appendix E: Health outcomes

We also estimated VA production functions for the height-for-age Z-score, a measure of health that is known to be sensitive to nutrition. The specification is identical to the pooled version of equation (9) that is discussed in Section 9.1 and reported in Table 6, column 1, except the health measures are now the dependent variables. Table E.1 presents the key results.

Table E.1. Height and weight (Z-scores) regressions (leisure omitted).

	Height-for-Age Z-Score		Weight-for-Age Z-Score	
Dependent Variable:	Beta	Std. Err.	Beta	Std. Err.
Lagged Z-score	0.6916	0.0146	0.8939	0.0122
Wealth Index	0.3735	0.0578	0.4203	0.0520
Brothers (#)	−0.0188	0.0068	−0.0443	0.0086
Sisters (#)	−0.0267	0.0057	−0.0300	0.0108
Mother: Primary educ	0.0626	0.0158	0.0524	0.0275
Mother: Secondary educ	0.0971	0.0252	0.1132	0.0393
Mother: Post-secondary educ	0.1581	0.0322	0.1469	0.0458
Market/Farm work (hrs/day)	−0.0017	0.0062	0.0126	0.0095
Chores (hrs/day)	0.0008	0.0045	0.0058	0.0093
School (hrs/day)	0.0013	0.0042	0.0086	0.0081
Study (hrs/day)	0.0058	0.0047	0.0045	0.0086
Sleep (hrs/day)	−0.0081	0.0050	0.0005	0.0083
Adjusted R^2	0.595		0.784	
Sample size	17,208		6446	
F-statistic for time allocation	1.428		0.489	
p-value	0.252		0.781	

Note: The Height-for-Age Z-score regression is pooled across ages 8 to 15 and four countries, with dummies added for each age and country combination in the sample. The Weight-for-Age Z-score is a regression on Age 8 individuals only, as the data is unavailable for other ages.

Our family resource measures are highly significant and with the expected signs. The wealth index and mother's education have strong positive effects on height.[3] Having more siblings reduces height as it means less resources are available for the target child.

Importantly, we find that the vector of child time-use is completely insignificant. This implies time use does not affect height conditional on resources. It also implies that time use is uncorrelated with nutrition inputs that are not captured by our resource measures. This is strong evidence that time use is also uncorrelated with unmeasured goods inputs into child development. We obtain similar results for the weight-for-age Z-score measured at age 8.

Appendix F: Lagged test scores and time inputs

We estimate "extended" value added models as discussed in Todd and Wolpin (2007). Two key features of our extended value-added models are that: (i) we control for lagged test scores and lagged time inputs, and (ii) we allow the effects of these lagged inputs (as well as the coefficients on current inputs and background variables) to differ by both age and country. Table F.1 reports the coefficients on lagged math scores in the math ability equations. Notice that the lagged math score, which the value-added model uses to control for unobserved ability, is highly significant in all instances. At ages 12, 15, and 19, the lagged score coefficients range from 0.352 to 0.722, with a median point estimate of 0.460.

[3]Interestingly, the education of the father was not significant (results available upon request).

Table F.1. Coefficients on lagged test score—MATH equation.

	Ethiopia		India		Peru		Vietnam	
	Beta	Std.Err.	Beta	Std.Err.	Beta	Std.Err.	Beta	Std.Err.
Age 8: Lagged math score age 5	0.049	0.014	0.150	0.018	0.103	0.016	0.091	0.021
Age 12: Lagged math score age 8	0.560	0.053	0.476	0.033	0.539	0.030	0.352	0.045
Age 15: Lagged math score age 12	0.360	0.037	0.394	0.031	0.398	0.034	0.394	0.037
Age 19: Lagged math score age 15	0.560	0.037	0.676	0.042	0.722	0.046	0.444	0.043

However, the coefficients on the age 5 score are much smaller in magnitude. As we note in Section 4.1, age 5 math skills were assessed using the CDA, an instrument designed for very young children, while math skills at later ages were assessed using paper and pencil tests. The changing nature of the test explains the smaller coefficient on lagged score at age 8.

Table F.2 reports the coefficients on lagged PPVT scores in the verbal ability equations. Again the lagged score is significant for all country/age combinations. The lagged score coefficients range from 0.218 to 0.625.

The lagged time input coefficients in the country/age specific models of Section 6 are too numerous to report. But we can highlight some key results. First, the five lagged time inputs are jointly significant (based on the F-test) in 13 of the 16 math equations. Thus, including lagged inputs (as suggested by Todd and Wolpin (2007)) clearly improves the fit of the math ability equations. The lagged time inputs are somewhat less important in the PPVT equations (i.e., they are jointly significant in only 6 of the 12 verbal ability equations).

In Section 9.1 of the main text, we present results where we pool the data across countries and ages. In this model the number of lag coefficients is small enough to report. These results are shown in Table F.3. The left side labeled "VA-OLS" reports the same model as in Table 6 column (1) of the main text, except here we also show the

Table F.2. Coefficients on lagged test score—PPVT equation.

	Ethiopia		India		Peru		Vietnam	
	Beta	Std.Err.	Beta	Std.Err.	Beta	Std.Err.	Beta	Std.Err.
Age 8: Lagged PPVT score age 5	0.221	0.051	0.263	0.040	0.363	0.038	0.218	0.028
Age 12: Lagged PPVT score, age 8	0.394	0.076	0.336	0.028	0.487	0.031	0.278	0.045
Age 15: Lagged PPVT score, age 12	0.285	0.072	0.430	0.060	0.625	0.040	0.279	0.053

Table F.3. Pooled math results for instrumenting the lag test score (study omitted).

	VA–OLS		VA–IV	
	Beta	S.E.	Beta	S.E.
Lagged Test Score	0.270	0.014	0.474	0.015
Market/Farm work (hrs/day)	−0.058	0.006	−0.049	0.004
Chores (hrs/day)	−0.058	0.006	−0.049	0.004
School (hrs/day)	−0.018	0.007	−0.014	0.004
Leisure (hrs/day)	−0.055	0.005	−0.046	0.004
Lag - Market/Farm work (hrs/day)	−0.051	0.007	−0.032	0.005
Lag - Chores (hrs/day)	−0.048	0.007	−0.031	0.005
Lag - School (hrs/day)	−0.018	0.008	−0.009	0.004
Lag - Leisure (hrs/day)	−0.036	0.007	−0.021	0.004

Note: The IV regression involves instrumenting the lagged math test score with the lagged verbal test score.

lagged time input coefficients. They have a very similar pattern to the current time input coefficients and are slightly smaller in magnitude.

As we discussed in Section 5, if test scores measure ability with error the coefficient on the lagged score will tend to be biased downward, causing the coefficients on time inputs to be biased in an ambiguous direction. To address this issue, we reestimated our models using the lagged PPVT score to instrument for the lagged math score, and vice versa. We find that these are strong instruments. The right side of Table F.3 labeled "VA-IV" shows the pooled model results. Instrumenting causes the lagged test score coefficient to increase from 0.27 to 0.47, as we would expect if there is independent measurement error in the two tests. The time use coefficients get slightly smaller but their pattern is unchanged.

We also redid our disaggregate VA models in the same way. When we instrument the lagged score coefficients increase (typically by 50% to 100% of the values reported in Tables F.1 and F.2). However, we find the coefficients on the time use variables are not much affected (results available on request).

Finally, another potential concern with our results is measurement error in the parent reports of child time use. But at ages 12 and 15 children were also asked about their own time use. Thus, in this subset of cases, we can use the child's reports to instrument for the parent's reports. While the IV estimates are slightly less precise, they are very similar to the OLS estimates (results available on request).

Appendix G: Effect of school relative to study and work relative to chores

In Figure G.1, we present results of the VA model for the effect of school time on test scores where study is the omitted category. This allows us to test the productivity of an additional hour of school relative to study. In the large majority of cases, the point estimates are negative, implying that time spent studying is more productive than an additional hour spent in school, but the difference is statistically insignificant in most cases. Combining the math and verbal results, we find that in only 3 out of the 28 cases is the difference significant at the 5% or 1% level.

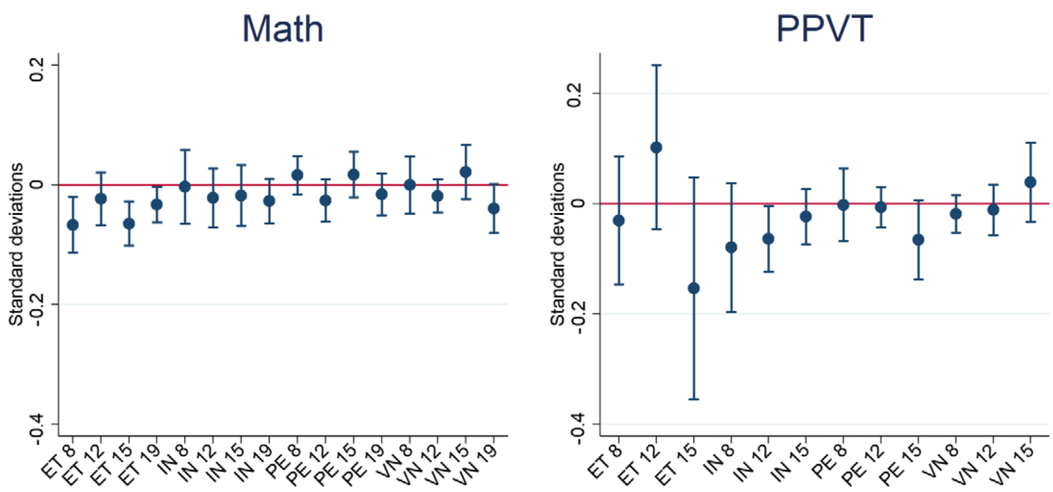


FIGURE G.1. Effect of schooling on math and verbal scores (relative to study).

In Figure G.2, we make domestic chores the omitted category and consider the effect of market/farm work relative to chores. 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 of 12 cases. This graph reinforces one of the main results of the article—that domestic chores is just as detrimental to a child's cognitive development as market/farm work when the counterfactual is clearly defined.

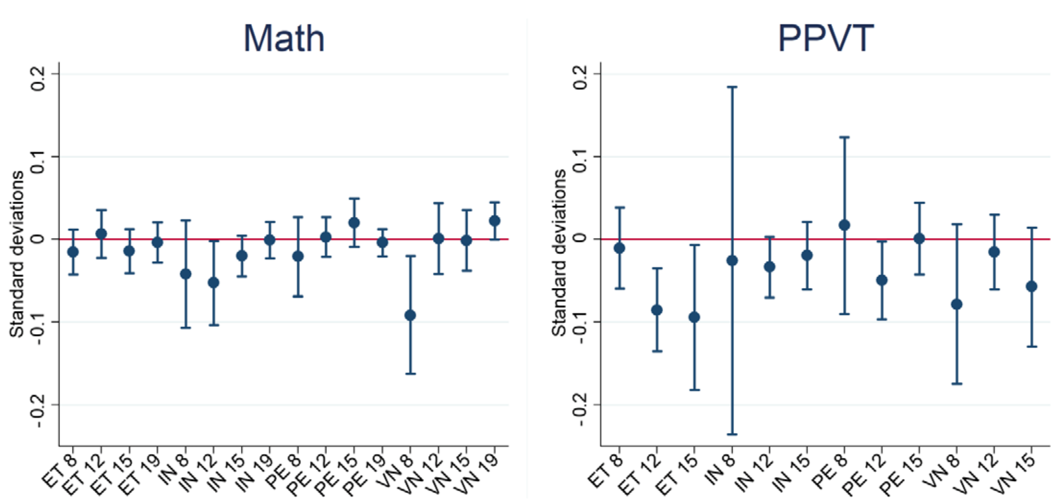
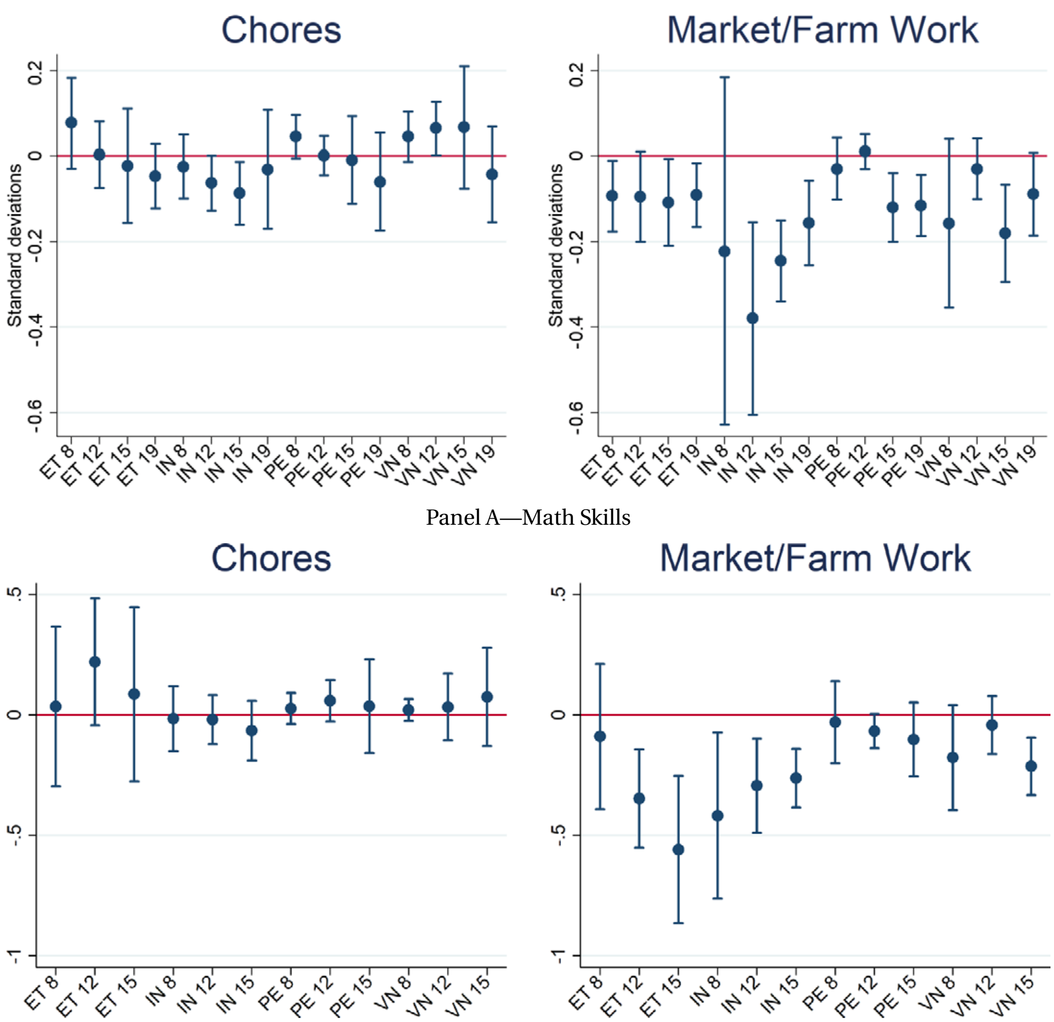


FIGURE G.2. Effect of market/farm work on math and verbal scores (relative to chores).

Appendix H: "Status quo" results using indicators for work and chores

Some "status quo" studies include an indicator for whether a child works, rather than actual work hours.[4] We replicate this type of analysis by reestimating equation (9) 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 H.1.



Panel A—Math Skills

Panel B—Verbal Skills

Figure H.1. "Status quo" analysis of impact of child work on math and verbal scores. *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.

[4] Examples of this type of analysis are Amin, Quayes, and Rives (2006), Beegle, Dehejia, and Gatti (2009), Emerson, Ponczek, and Portela Souza (2017), Patrinos and Psacharopoulos (1997), Psacharopoulos (1997), Ravallion and Wodon (2000).

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.

Appendix I: Assessing the role of unobservables

Here, we assess the role of unobservables in our results. In our VA model, we control for unobserved ability using the 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) measures of household resources available to the child, including parental education, household wealth, number of siblings, and height-for-age, and (3) several additional controls that serve as further proxies for past inputs and child ability.[5] 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 I.1 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 -0.143 for chores and -0.137 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 -0.083 for chores and -0.104 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 -0.048 for chores and -0.055 for market/farm work. Nevertheless, the estimates remain highly significantly negative in 11 out of 16 cases for chores and 14 out of 16 cases for market/farm work.

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 direction of 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.

Appendix J: Details of the construction of instruments

Here, we list all the instruments included in the 2SLS models in Tables 4 and 5 in the paper, the methodology used to construct them, and the instruments that were picked

[5] As described in Section 4, these include child age, gender, religion, ethnicity, parental age, and whether in household, and whether household is urban/rural, household member composition.

Table I.1. 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	−0.096	−0.092	−0.182	−0.108	−0.107
	(0.034)	(0.021)	(0.021)	(0.032)	(0.017)	(0.017)
India	−0.179	−0.092	−0.083	−0.140	−0.129	−0.125
	(0.037)	(0.034)	(0.032)	(0.052)	(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	−0.082	−0.057	−0.154	−0.152	−0.110
	(0.036)	(0.026)	(0.020)	(0.030)	(0.029)	(0.027)
Peru	−0.113	−0.039	−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	−0.071	−0.227	−0.099	−0.070
	(0.028)	(0.021)	(0.021)	(0.036)	(0.012)	(0.010)
Age 15						
Ethiopia	−0.187	−0.098	−0.067	−0.160	−0.109	−0.082
	(0.022)	(0.017)	(0.014)	(0.026)	(0.018)	(0.017)
India	−0.130	−0.067	−0.035	−0.129	−0.100	−0.055
	(0.031)	(0.026)	(0.023)	(0.031)	(0.023)	(0.020)
Peru	−0.119	−0.070	−0.039	−0.101	−0.055	−0.019
	(0.025)	(0.018)	(0.015)	(0.024)	(0.015)	(0.015)
Vietnam	−0.110	−0.041	−0.025	−0.120	−0.042	−0.026
	(0.026)	(0.018)	(0.016)	(0.027)	(0.017)	(0.014)
Age 19						
Ethiopia	−0.151	−0.103	−0.041	−0.120	−0.108	−0.045
	(0.020)	(0.015)	(0.015)	(0.014)	(0.012)	(0.010)
India	−0.137	−0.102	−0.054	−0.134	−0.116	−0.055
	(0.016)	(0.016)	(0.012)	(0.019)	(0.017)	(0.015)
Peru	−0.046	−0.046	−0.027	−0.050	−0.051	−0.031
	(0.014)	(0.014)	(0.015)	(0.017)	(0.016)	(0.015)
Vietnam	−0.142	−0.101	−0.073	−0.122	−0.090	−0.051
	(0.030)	(0.023)	(0.019)	(0.020)	(0.017)	(0.017)

Note: Omitted time category is study in all regressions. Standard errors are in parentheses.

by LASSO for the reduced instrument set. There are two types of instruments we considered: instruments that vary across communities and those that vary across households. Several of the household-level instruments and controls were also interacted with the community-level instruments and rainfall to capture additional variation in the first stage regression.

Table J.1 details the community-level instruments, which are principal components of community prices, wages, and services. Due to changes in the questions that were

Table J.1. Community-level instruments.

Instrument	Methodology
Community Prices	Young Lives conducts community-level surveys with a representative from each community. The two instruments used here are the first two principal components of prices for the following goods: Grade 3 textbook, notebook, shoes, boy's shirt, girl's shirt, boy's pants, girl's skirt, oral rehydration salts, paracetamol, amoxicillin, mebendazol, cigarettes, detergent, kerosene, fertilizer – dap, fertilizer – urea, barley, wheat, corn, sorghum, coffee, sugar, salt, white teff, black teff, millet, oats, horse beans, cow peas, chick peas, field peas, other pulse, milk, yogurt, butter, eggs, beef, mutton, chicken meat, goat meat, other meat, chat, araqi, tej, cooking oil, karia, berbere, other spices, bread, enset, potato, gommen selata, jinjibel, tikl gommen, nech shinkuri, fasolia, fenugreek, onions, and vegetable.
Community Wages	Young Lives conducts community-level surveys with a representative from each community. The two instruments used here are the first two principal components of the following average daily wage information for males and females (when available): Age 8 (Round 3): Prepare the land for agricultural use, plant/sow, weed out agricultural land, harvest, perform post-harvest duties, to pasture/to put cattle to graze, shepherding, helping at workshops, construction worker, factory worker, taxi/minibus driver, security guard, maid/domestic worker, shop assistant, teacher, police, mechanic, cook, fisherman, tailor, military, computer operator, and other tasks. Age 12 (Round 4): Prepare the land for agricultural use, plant/sow, weed out agricultural land, harvest, perform post-harvest duties, to pasture/to put cattle to graze, shepherding, and other tasks.
Community Services	The 'service/infrastructure factors' are the first two principal components derived from a range of variables in the community-level surveys that seek to measure the range and quality of infrastructure or services in the community. The variables are defined as S_{ic} for type of infrastructure/service i and community c. $S_{ic} = 0$ if the type of infrastructure i is absent from the community, $S_{ic} = 1$ if i is present but considered "bad" in quality, $S_{ic} = 2$ if i is present and considered "so-so" in quality, and finally $S_{ic} = 3$ if present and considered "good" in quality. The infrastructure or services included are as follows: Age 8 and age 12: space exclusively assigned for little children (i.e., children's playground), sporting fields, camping zones or family recreational areas, indoor/outdoor movie theatres, video games, public telephones, private telephones, public internet cabin, electricity, drinkable water, sewerage, police station, public bank, and private bank. Age 8 only: fairgrounds for temporary recreational activities, religious institutions (church, mosque, etc.), and schools. Age 12 only: private internet cabin, mobile phone service, primary agricultural cooperative society, and local government credit/savings schemes.

asked between Rounds 3 and 4 of the Young Lives surveys, the components of each category may change between the IV regression for age 8 (which is Round 3) and age 12 (which is Round 4).

Table J.2 details the household-level instruments. Several of them have been interacted with the community-level factors, as well as the rainfall instruments, to capture

Table J.2. Household-level instruments.

Instrument	Methodology
Agricultural prices (as levels and interacted with gender)	$$AP_{it} = \sum_{c=1}^{C} w_{it,c} R_{it.c} / \sum w_{it,c}$$ where $w_{it,c}$ is the quantity of crop c that is grown by individual i's family at time t (if the household does not grow crops, then $w_{it,c} = 0\ \forall c$ and $AP_{it} = 0$). $R_{it,c}$ is the relative price of crop c at time t, defined as the price of crop c at time t for household i (which varies by community) divided by the average price for the same price measured in the same community at all observed time periods. The crops included in the set C are: barley, black/mixed teff, white teff, cabbage, chat, chick peas, cow peas, wheat, sorghum, potatoes, onions, oats, maize, horse beans, vegetables, and coffee.
Livestock prices (as levels and interacted with gender)	$$LP_{it} = \sum_{l=1}^{L} w_{it,l} R_{it.l} / \sum w_{it,l}$$ where $w_{it,l}$ is the value of livestock l that is held by individual i's family at time t (if the household does not hold livestock, then $w_{it,l} = 0\ \forall l$ and $\mathrm{L}P_{it} = 0$). $R_{it,l}$ is the relative price of livestock l at time t, defined as the price of livestock l's output at time t for household i (which varies by community) divided by the average price for the same output measured in the same community at all observed time periods. The livestock considered in this analysis, and the definition of its output, is as follows: The output price for cows is the average price of milk, beef, and yogurt. The output price for chickens is the average price of chicken meat and eggs, the output price for sheep is the price of mutton, and the output price for goats is the price of goat meat.
Whether the father is employed in agriculture	Constructed dummy variable from Young Lives survey data. It equals 1 if child father's "most important activity" is reported as either: "self-employed (food crops)," "self-employed (aquaculture)," "self-employed (livestock)," "wage employment (agriculture)," "annual farm servant'," or "other agricultural work," and 0 otherwise.
Month of interview dummies	Young Lives provides date of interview in DD/MM/YY format. The month of the interview is extracted from the date and turned into a series of dummies. For age 8 the following month dummies are used: 1, 10, 11, 12. For age 12, we use dummies for month: 1, 2, 10, 11, 12.
Environmental Shock–Frost	The Young Lives survey asks respondents if they had experienced frosts.
Time to school	Young Lives survey question which asks respondent to report 'Travel time to school (in minutes)'.
Rainfall in last 2 and 12 months	*Step1: Obtain rainfall raster images at the relevant dates* Satellite imagery of rainfall was obtained using the Tropical Rainfall Measuring Mission (TRMM), a joint mission by NASA and the Japanese Space Agency (JAXA). The rainfall maps are at 0.25 degrees resolution and collected for all months around the date of interviews in Rounds 3 and 4 of the Ethiopia Young Lives Surveys. Available online at: https://neo.sci.gsfc.nasa.gov/view.php?datasetId=TRMM_3B43M

(*Continues*)

Table J.2. *Continued.*

Instrument	Methodology
	Step 2: Estimate the GPS location of Ethiopian communities in the Young Lives surveys The GPS locations of the communities surveyed in Young Lives were estimated from the sentinel site information provided by Table 1 of the Young Lives Survey Design and Sampling (Round 5) document for Ethiopia. Available online at: https://www.younglives.org.uk/content/survey-design-and-sampling-round-5-ethiopia *Step 3: Construct the instruments* The two instruments measure the demeaned rainfall in the last 2 and 12 months. The monthly sum of precipitation at the child's community GPS location is aggregated for 2/12 months prior to the household-level date of interview (usually interviews in a single community is conducted over 2 months, so there is a little variation within communities). This sum of rainfall over the last 2/12 months is then demeaned using the average rainfall over those same months over 2000–2019. For example, if the household was interviewed in October 2008, the rainfall in the last 2 months instrument is calculated from: Rainfall at GPS location in October 2008 + Rainfall at GPS location in September 2008—average rainfall at GPS location in October 2000–2019— average rainfall at GPS location in September 2000–2019.

additional variation. These are time to school, livestock prices, agricultural prices, frost, and the month of interview dummies. Several control variables were also interacted, which are gender, the wealth index, the urban dummy, the number of brothers in the household, and the number of sisters in the household. Lastly, the community-level factors and rainfall were also interacted with the number of older siblings in the household.

To select the subset of instruments that best explain the independent variation of each endogenous regressor, we adopt LASSO with regularization parameter λ. To find the optimal λ for each endogenous regressor, we estimate LASSO across a grid of λ val-

Table J.3. Instruments chosen by LASSO.

Regression	Instruments Chosen by LASSO in the First Stage Regressions
Age 8 (no lagged test score)	*Market/farm work*: Agricultural Prices * Gender, Time to school, Father in Agriculture, Rain (last 2 months) * Agricultural Prices, Wage factor 1 * Time to School, Wage factor 1 * Livestock Prices, Price factor 1 * Gender, Price factor 1 * # of Older Siblings *Chores*: Rain (last 2 months) * Urban, Wage factor 1 * Agricultural Prices, Price factor 1 * Urban, Services factor 1 * # of Sisters, Services factor 1 * Agricultural Prices, Services factor 2 * Time to School *Leisure*: Rain (last 12 months), Rain (last 12 months) * Agricultural Prices, Wage factor 2 * Time to School, Wage factor 2 * # of Older Siblings, Price factor 2 * Gender, Price factor 2 * Urban *Sleep*: Services factor 2, Wage factor 2 * October dummy, Price factor 2 * Frost Shock, Services factor 2 * Agricultural Prices

(*Continues*)

Table J.3. *Continued.*

Regression	Instruments Chosen by LASSO in the First Stage Regressions
Age 8 (with lagged test score)	*Market/farm work*: Agricultural Prices * Gender, Time to School, Father in Agriculture, Rain (last 2 months) * Agricultural Prices, Wage factor 1 * Time to School, Wage factor 1 * Livestock Prices, Price factor 1 * Gender *Chores*: Services factor 2, Price factor 1 * Urban, Price factor 1 * # of Older Siblings, Services factor 2 * Time to School *Leisure*: Rain (last 12 months) * Agricultural Prices, Wage factor 2 * Time to School, Wage factor 2 * # of Older Siblings, Price factor 2 * Gender, Price factor 2 * Urban *Sleep*: Rain (last 12 months) * November Dummy, Price factor 2, Price factor 2 *Frost Shock, Services factor 1 * # of Brothers
Math Test Scores: Age 12 (no lagged test score)	*Market/farm work*: Livestock Prices * Gender, Rain (last 12 months) * Livestock Prices, Services factor 2 * Gender, Services factor 2 * Frost Shock *Chores*: Wage factor 1, Wage factor 2 * Gender, Price factor 2 * Urban, Services factor 1 * Wealth index, Services factor 2 * Wealth index, Services factor 2 * Urban *Leisure*: Rainfall (last 2 months), Rainfall (last 2 months) * Urban, Rainfall (last 12 months) * Wealth index, Rainfall (last 2 months) * Gender, Wage factor 2 * Urban, Services Factor 1 * November dummy, Services Factor 2 * November dummy *Sleep*: Wage factor 2, Wage factor 2 * Frost Shock
Math Test Scores: Age 12 (with lagged test score)	*Market/farm work*: Livestock Prices * Gender, Rain (last 12 months) * Livestock Prices, Services factor 2 * Gender, Services factor 2 * Frost Shock, Services factor 2 * # of Older Siblings *Chores*: Wage factor 1, Wage factor 2 * Gender, Price factor 2 * Urban, Services factor 1 * Wealth index, Services factor 2 * Wealth index, Services factor 2 * Urban *Leisure*: Rainfall (last 12 months), Rainfall (last 2 months) * Urban, Rainfall (last 12 months) * Gender, Rainfall (last 12 months) * Wealth index, Wage factor 1 * November dummy, Wage factor 2 * Urban, Services factor 1 * November dummy *Sleep*: Wage factor 2, Wage factor 2 * Frost dummy
Verbal Test Scores: Age 12 (no lagged test score)	*Market/farm work*: Livestock Prices * Gender, Rain (last 12 months) * Livestock Prices, Services factor 1 * Gender, Services factor 1 * Wealth Index, Services factor 2 * Frost Shock, Services factor 2 * # of Older Siblings *Chores*: Wage factor 1, Services factor 2 * Urban *Leisure*: Rainfall (last 12 months), Rainfall (last 2 months) * Urban, Rainfall (last 12 months) * Gender, Wage factor 2 * Urban, Services factor 1 * November dummy, Services factor 2 * Wealth Index, Services factor 2 * Livestock prices *Sleep*: Wage factor 2, Wage factor 2 * Frost Shock, Wage factor 2 * October dummy
Verbal Test Scores: Age 12 (with lagged test score)	*Market/farm work*: Livestock Prices * Gender, Rain (last 12 months) * Livestock Prices, Services factor 2 * Gender, Services factor 2 * Frost Shock, Services factor 2 * # of Older Siblings *Chores*: Wage factor 1, Wage factor 2 * Gender, Price factor 2 * Urban, Services factor 1 * Wealth index, Services factor 2 * Wealth index, Services factor 2 * Urban *Leisure*: Rainfall (last 12 months), Rainfall (last 2 months) * Urban, Rainfall (last 12 months) * Wealth index, Rainfall (last 12 months) * Gender, Wage factor 2 * November dummy, Wage factor 2 * Urban, Services factor 1 * November dummy *Sleep*: Wage factor 2, Wage factor 2 * Frost Shock

ues and use the Bayesian Information Criterion (BIC) to find the optimal value. The grid was constructed with a maximum of $2 \cdot \max(Z_i' x_i)$, where Z is the instrument vector and x the endogenous regressor in question, and a minimum value that is 1/1000 the max-

Table J.4. Sample sizes for Tables 4 and 5.

	Math Age 8	Math Age 12	Verbal Age 8	Verbal Age 12
OLS (No Controls)	1824	1869	1602	1654
OLS	1818	1863	1596	1648
VA	1240	1799	1031	1572
IV (Lagged Test Excluded):				
*2*SLS	1686	1673	1471	1465
*2*SLS (Lasso)	1686	1673	1471	1465
*G*MM (Lasso)	1686	1673	1471	1465
IV (Lagged Test Included):				
*2*SLS	1654	1615	1371	1396
*2*SLS (Lasso)	1654	1615	1371	1396
*G*MM (Lasso)	1654	1615	1371	1396

Note: In contrast to the main VA results in the paper, all language groups were included in the verbal test score regressions here, resulting in a significantly larger sample size. This was done to ensure IV would have a sufficiently large sample for estimation and inference.

imum. This follows typical practice in the literature. The number of grid points was set to 25 for Ethiopia, age 8 and 500 for age 12, as we went with the grid that gave the best first-stage F results.

Table J.3 lists the instruments that were picked by LASSO. This differs by age group and by whether the lagged test score was included in the IV regression as a control variable. It also differs slightly between the verbal test score regressions at age 12 due to the different sample size available for instrument selection. Lastly, Table J.4 lists the sample sizes used for Tables 4 and 5 in the main body of the paper.

Appendix K: Estimated effects of household resources on test scores

Table K.1 presents results of the household resource variables included in the pooled regression of Table 6, column 1 but not shown for brevity. It lists results for both math scores and verbal scores, with study as the omitted category.

Many of the variables are highly significant for both math and verbal scores, with wealth, nutrition (height-for-age Z-score), family composition (e.g., number of siblings and whether both parents are present), and parental education all strongly affecting the cognitive development of the child in the expected ways.

Interestingly, unlike the health outcome regressions, where only mother's education is significant (see Appendix E), both parents' education is important for test scores.

Appendix L: Nonlinearities in time allocation effects

In this section, we examine nonlinearities in the effect of time use. To do this, we combine school and study into a single time use category. Table L.1 presents results from models that pool across all countries and age groups, and that allow for nonlinearity in the effects of time use. Leisure is the omitted time category in all cases.

Table K.1. Pooled math and verbal score regressions (study omitted).

Dependent Variable:	Maths Score Beta	Maths Score Std. Err.	Verbal Score Beta	Verbal Score Std. Err.
Lagged Score	0.270	0.014	0.300	0.018
Wealth Index	0.273	0.039	0.568	0.081
Height-for-age Z-score	0.043	0.005	0.063	0.010
Brothers (#)	−0.011	0.004	−0.034	0.008
Sisters (#)	−0.006	0.004	−0.030	0.007
Mother Ed: Primary	0.069	0.016	0.106	0.024
Mother Ed: Secondary	0.118	0.019	0.161	0.031
Mother Ed: Post-secondary	0.180	0.021	0.247	0.040
Father Ed: Primary	0.080	0.017	0.081	0.028
Father Ed: Secondary	0.126	0.019	0.102	0.029
Father Ed: Post-secondary	0.173	0.022	0.178	0.033
Both Parents in household	0.024	0.013	0.023	0.021
Urban dummy	0.042	0.026	0.194	0.064
Adjusted R^2	0.704		0.682	
Sample size	20,330		13,841	

Note: Table 6 column 1 in the main text reports the time use coefficients from the pooled math regression, and Table F.3 reports the lagged time use coefficients.

The new baseline model in Table L.1, column (1) reproduces the main result from Table 6, column 1 that market/farm work and chores are not detrimental to cognitive development relative to leisure. But they are detrimental if they crowd out school/study.

Model (2) adds a quadratic term for school/study hours, while model (3) does the same for market/farm work and chores separately and model (4) adds all three. A slight concavity can be observed in the relationship between school/study hours and test scores in the first and third model, but it is statistically insignificant. Quadratic terms in hours of work and chores are not significant in columns (3) and (4) either. Also, the estimated effects of school and work time are little affected by the inclusion of the quadratic terms. Thus, our main conclusion is not sensitive to a more flexible specification of the effects of time use.

Appendix M: Dynamic complementarities in time allocation

Here, we examine whether effects of time use differ by skill level. To this end, we combine school/study into one category, and interact time use with the lagged math ability test score. The results in Table M.1. Column (1), which combines school/study without including interactions, confirm our previous result that school/study is beneficial relative to all alternative times uses, while work, chores, and leisure are equivalent. Column (2) interacts school/study time with the lagged math ability test. Interestingly, the interaction is significant and *negative*. This implies additional school time would be more beneficial for relatively low ability children. In column (3), we interact all time use variables with the lagged math score. The results imply that child work is less detrimental

TABLE L.1. Pooled math results with interactions (leisure omitted).

	(1)	(2)	(3)	(4)
Lagged Test Score	0.271	0.271	0.271	0.271
	(0.014)	(0.014)	(0.014)	(0.014)
School/Study (hrs/day)	0.043	0.051	0.043	0.052
	(0.003)	(0.008)	(0.003)	(0.008)
Market/Farm Work (hrs/day)	−0.002	−0.001	−0.007	−0.007
	(0.003)	(0.003)	(0.006)	(0.006)
Chores (hrs/day)	−0.003	−0.002	−0.001	−0.003
	(0.004)	(0.004)	(0.006)	(0.006)
(School/Study hrs)ˆ 2/100		−0.072		−0.079
		(0.053)		(0.055)
(Work hrs)ˆ 2/100			0.054	0.069
			(0.059)	(0.061)
(Chores hrs)ˆ 2/100			−0.015	0.008
			(0.061)	(0.063)
R^2	0.704	0.704	0.704	0.704
N	20,330	20,330	20,330	20,330

Note: Omitted time category is leisure in all regressions. Standard errors are in parentheses.

(relative to school) for higher ability children. These results are consistent with a model where school and ability are substitutes in the production of cognitive ability, while ability complements learning from work and chores.

TABLE M.1. Pooled math results with interactions (leisure omitted).

	(1)	(2)	(3)
Lagged Test Score	0.271	0.362	0.232
	(0.014)	(0.023)	(0.028)
School/Study (hrs/day)	0.043	0.042	0.043
	(0.003)	(0.003)	(0.003)
Market/Farm Work (hrs/day)	−0.002	−0.003	−0.001
	(0.003)	(0.003)	(0.003)
Chores (hrs/day)	−0.003	−0.003	0.000
	(0.004)	(0.004)	(0.004)
Lagged Test * School/Study		−0.012	−0.001
		(0.002)	(0.002)
Lagged Test * Market/farm work			0.017
			(0.003)
Lagged Test * Chores			0.017
			(0.004)
R^2	0.704	0.705	0.707
N	20,330	20,330	20,330

Note: Omitted time category is leisure in all regressions. Standard errors are in parentheses.

References

Amin, Shahina, Shakil Quayes, and Janet M. Rives (2006), "Market work and household work as deterrents to schooling in Bangladesh." *World Development*, 34 (7), 1271–1286. [12]

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), 871–889. [12]

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), 175–187. [3]

Emerson, Patrick M., Vladimir Ponczek, and André Portela Souza (2017), "Child labor and learning." *Economic Development and Cultural Change*, 65 (2), 265–296. [12]

Patrinos, Harry A. and George Psacharopoulos (1997), "Family size, schooling and child labor in Peru—an empirical analysis." *Journal of Population Economics*, 10 (4), 387–405. [12]

Psacharopoulos, George (1997), "Child labor versus educational attainment some evidence from Latin America." *Journal of Population Economics*, 10 (4), 377–386. [12]

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–C175. [12]

Royston, P. and D. G. Altman (1994), "Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling." *Applied Statistics*, 43, 429–467. [5]

Royston, P. and W. Sauerbrei (2008), *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Wiley, Chichester, UK. [5]

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. [8, 9]

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* (Wim J. van der Linden and Ronald K. Hambleton, eds.). Springer, New York, N.Y. [3]

Verriest, Ewout (2020), "Household labor supply, child development and childcare policies in the United States." SSRN Research Paper 3330051. Available Online: http://dx.doi.org/10.2139/ssrn.3330051. [3]

Co-editor Limor Golan handled this manuscript.

Manuscript received 14 October, 2020; final version accepted 18 October, 2021; available online 3 December, 2021.