

Childhood Learning and Adult Earnings in Four LMIC Cohorts*

Jishnu Das Yash Dhuldhoya Marta Favara Amory Gethin
Alan Sánchez Abhijeet Singh

October 16, 2026

Abstract

Researchers and policymakers in low- and middle-income countries (LMICs) have customarily relied on indirect evidence to justify investments in learning, by either linking test scores to economic growth or connecting adult cognitive assessments to earnings. The reliance on indirect evidence is largely because longitudinal data linking school-age learning to adult earnings are scarce. We address this gap by using data from Ethiopia, India, Pakistan, and Peru to estimate associations between foundational learning (FL), measured by test scores around age 12, and human capital and earnings in young adulthood. In Peru, India, and Pakistan, a one-standard-deviation increase in FL is associated with an 8–13 log point increase in adult earnings. Childhood scores also predict 1.1–1.5 additional years of schooling and 9.3–13.5 percentage points higher college or vocational attendance, as well as higher skills in adulthood. Interestingly, a linear-projection decomposition indicates that measured schooling and adult test scores span from 18% in Peru to 50–57% of the FL-earnings gradient in India and Pakistan, suggesting that FL may raise earnings over and above the association with schooling and adult scores. Finally, FL can account for about 8–22 percent of the earnings gap between young Ethiopians and Indians in our sample, a number that could rise to 50 percent as returns to learning grow over the life cycle. We interpret these descriptive findings in light of recent causal evidence on the learning-to-earnings pipeline and discuss the implications for policy investments in FL.

JEL codes: O12, I21, I23, J24

Keywords: learning, education, schooling, rate of return, human capital

**Affiliations:* Das: Georgetown University, NBER, and CGD; Dhuldhoya: Georgetown University; Favara: Young Lives, University of Oxford; Sánchez: Economics Department, Leeds University Business School, University of Leeds; Gethin: World Bank; Singh: Stockholm School of Economics, CEPR, and BREAD.

Acknowledgments: This paper is prepared in honour of Prof. Jere Behrman who did pioneering work on child development and its consequences in LMICs, including all our study countries, and supported long-term longitudinal studies globally. Favara, Sanchez and Singh particularly thank him for his personal mentorship, and his long-standing support to the Young Lives study. We thank Lant Pritchett for his insightful comments, and Tanimata Tanimata and Amanda Woodman Deza for their support in analyzing the Young Lives data collected in 2023/24. Das thanks the What Works Hub for Global Education, the Gates Foundation and the Marshall Foundation for support. Gethin thanks the UK Government through the Data and Evidence for Tackling Extreme Poverty (DEEP) Research Program for support. Favara thanks the UK Foreign, Commonwealth & Development Office (FCDO) under Grant [number VGR01421], Irish Aid under Grant [number VGR01420], and Echidna Giving under Grant [number 11414775] for support. Singh thanks the Handelsbanken Research Foundations and the What Works Hub for Global Education for support.

1 Introduction

The economic rationale for investing in foundational learning (hereafter, FL) in primary schools rests on a straightforward calculation: By how much does a program that strengthens FL increase future earnings and possibly other indicators of well-being? Direct evidence of this link, however, remains scarce in low- and middle-income countries (LMICs) because high-quality longitudinal datasets linking test scores at primary school age to adult earnings have been unavailable. This is a substantial omission since panel estimates of the wage return to test scores are central to comparative analyses of the effectiveness of public spending in different educational interventions (e.g. [Hendren and Sprung-Keyser, 2020](#)).¹

In this paper, we address this gap for four LMICs—Ethiopia, India, Pakistan, and Peru—using unique panel data that measure student test scores around age 12, earnings in early adulthood (~ 25 -29 years), and demographic characteristics. For Ethiopia, India (Andhra Pradesh and Telangana states), and Peru, the data come from the Young Lives study, which surveyed two cohorts of children from 2002 to 2023/24; and, for Pakistan, from the LEAPS study which tracked children from primary school (beginning in 2003) to 2018. Both studies included rich measures of test scores at multiple points in time, followed individuals who migrated to other locations and out of the initial (parental) household, and maintained low attrition. We use these data to examine how FL, measured as test scores when children are about 12 years old, relates to later-life educational attainment (years of schooling, high-school completion and college/vocational attendance) and, ultimately, to employment and earnings.

There are four main results from our analysis. First, in every country, FL is positively associated with years of schooling and college/vocational attendance.² Reflecting substantial gains in recent years, average years of schooling range from 9.4 in Pakistan to 13 in Peru, while college or vocational education attendance varies from 16 to 68 percent. A one-standard-deviation (σ) increase in FL is associated with 9.3 percentage points higher college or vocational education attendance in Pakistan, equivalent to 58 percent of the sample mean. Associations in the other countries are of comparable magnitude.

¹This omission is also in marked contrast to OECD economies where high-quality panel datasets have long been available (e.g. the National Longitudinal Survey of Youth in the US) and, more recently, linked administrative registers have allowed population-wide studies with well-powered investigations of causal effects of teacher-level or class-level shocks (see, e.g., [Chetty et al., 2014](#); [Fredriksson et al., 2013](#)). In practice, given the absence of similar data, the argument for investing in FL in LMICs has relied on two indirect strands of evidence: one that associates national literacy and numeracy levels with economic growth at the macro level (e.g. [Hanushek and Woessmann, 2008](#)), and another that connects adult skills to individual earnings at the micro level (e.g. [Valerio et al., 2016](#)).

²We have previously documented the FL–education relationship using earlier waves of the same datasets; the current estimates now draw on nearly completed schooling histories, enabling a more accurate assessment of the association between FL and college/vocational attendance ([Das et al., 2022](#)).

Second, in the subsample of individuals with positive earnings, a 1σ increase in FL is associated with 7.6 log points larger earnings in Pakistan and almost 13 log points larger earnings in Peru and India. Over the range of scores observed in the Indian and Peruvian samples, monthly earnings are nearly doubled. Log earnings increase linearly with test scores across the full range of achievement, and do so for both men and women, although they are imprecisely measured for women.³ In Ethiopia, we do not detect any statistically significant link between FL and young adult earnings. We interpret this result in light of the timing of the survey, conducted just as Ethiopia was emerging from an internecine civil war that resulted in conservative estimates of approximately 100,000 deaths (McGowan et al., 2025).

Third, the estimated associations between test scores at age 12 and earnings in adulthood are meaningfully larger than would be predicted solely through measured adult skills and years of schooling. We apply a predictive decomposition framework that uses the correlations between (i) childhood test scores and years of schooling, (ii) childhood test scores and adult test scores, (iii) earnings in adulthood and these two markers of human capital, and (iv) the covariance between these two markers of human capital. In the framework of Athey et al. (2025), this assesses whether years of schooling and measured cognitive skills in adulthood are adequate surrogate measures to capture the full association between childhood test scores and adult earnings. We find that they are not. In the sample of men with positive earnings, increases associated with these surrogates explain about a fifth of the measured FL-earnings gradient in Peru, half in Pakistan, and 57% in India.⁴

Fourth, we show in a simple growth accounting exercise that schooling and test scores combined explain a meaningful portion of the cross-sectional earnings differences between countries. Comparing Ethiopia and India, schooling and learning account for about 8-22% of the cross-sectional earnings difference among young adults. This number could rise to 50% after accounting for the fact that returns to schooling and learning tend to grow over the life cycle. Of this contribution, about 40-50% comes from differences in FL, while 50-60% comes from schooling. These estimates are likely a conservative lower bound given potential measurement error and human capital externalities that we do not incorporate in this analysis.

³There are substantial gender differences in the probability of having worked in all countries: about 80% of men in our samples in Ethiopia and Pakistan, and over 90% in India and Peru, report having worked, compared with about 7.5% of women in Pakistan, 39% in Ethiopia, 55% in India, and 71% in Peru.

⁴One potential explanation for the ability of FL to predict earnings beyond schooling and adult scores is a standard omitted variables bias. While this is certainly plausible—and our correlations should not be given a causal interpretation—there is also a reasonable argument that our estimates are a lower bound on the true effect of improving FL. This is because: (a) test scores are measured with error, which generates the usual attenuation bias; (b) we focus on early-career rather than lifetime earnings, even though age-earnings profiles vary by FL and educational attainment; and (c) the individuals for whom higher FL is likely to have the greatest effects are sometimes still in college and have not yet entered the labor market.

The principal contribution of our paper is to present and interpret the first multi-country exploration of the association between child test scores and adult labor market outcomes in low- and middle-income countries. Two recent papers provide the closest similar evidence. First, [Glewwe et al. \(2022\)](#) follow 2,000 children aged 9–12 in Gansu, China, starting in 2000 and re-interviewing them in 2015, when they are 24–27 years old. They find that a 1σ increase in test scores is linked to a 12.9 percent increase in earnings. Second, in work conducted in parallel to ours, [Crawford \(2026\)](#) analyzes the relationship between FL and earnings using the Indonesian Family Life Survey and likewise finds that a 1σ increase in test scores is associated with a 9 percent increase in earnings. Our paper provides evidence on four additional countries that span a wide range of development outcomes.⁵ Taken together, these estimates now offer an emerging reference range for linking FL and earnings later in life, which can serve as a basis for calculating the costs and benefits of alternative educational interventions.⁶

These results relate to several streams of research in development economics. Foremost, they provide the missing link between a substantial literature that evaluates interventions to increase test scores in primary schooling (see, e.g., [Angrist et al. \(2025\)](#)) and studies that link adult test scores to economic outcomes (see, e.g., [Chua \(2017\)](#); [Valerio et al. \(2016\)](#); [Ozawa et al. \(2022\)](#)). Our predictive decomposition also helps formalize the extent to which the association between child test scores and adult outcomes is greater than would be predicted solely through years of schooling and adult test scores, the two measures typically used in the literature to proxy for returns to educational investments earlier in childhood ([Ozawa et al., 2022](#)). Finally, we also provide a link between the micro-econometric literature evaluating interventions that raise test scores and the macroeconomic literature on growth accounting that has attempted, using several complementary data sources and strategies, to account for the contribution of human capital to cross-country income differences. A substantial theme in recent work in this area has been to broaden the definition of human capital from only years of schooling (e.g., [Mankiw et al. \(1992\)](#), [Caselli \(2005\)](#)) to also incorporate measures that account for quality differences (e.g. [Schoellman \(2012\)](#); [Manuelli and Seshadri \(2014\)](#); [Hendricks and Schoellman \(2018\)](#)). In looking at the association of childhood test scores with *both* years of schooling and adult test scores, and the further downstream association with adult earnings, we contribute to this discussion directly for a wide span of countries.

⁵The data used in this paper also have some further advantages including, especially, continuing to track migrants from original communities and individuals who form or move to other households. As shown by [Beegle et al. \(2011\)](#), tracking migrants in panel surveys is crucial for estimating economic mobility in LMICs.

⁶In this context, the null estimates from Ethiopia, though observed under the highly specific conditions of widespread conflict, serve as an important reminder that these relationships vary by setting, and that explaining this heterogeneity will require a significant effort in building more longitudinal datasets with a focus on causality.

2 Data and Context

2.1 Countries

The four countries we study span a range of incomes: Ethiopia is a low-income country (with GDP per capita PPP \$3,288 as of 2024); Pakistan and India are both lower-middle-income countries (\$6,251 and \$11,160, respectively); and Peru recently moved to upper-middle-income status (\$17,802). Recent estimates of “learning poverty” (World Bank et al., 2022) show that among the four countries, the lowest levels of reading proficiency by the end of primary are in Ethiopia and Pakistan, followed by India and Peru. Even in Peru, however, achievement remains low by international standards; it ranks near the bottom of the PISA rankings.

2.2 Analytic Samples

We use data from two cohort studies in this paper: the Learning and Education Achievement in Punjab Schools (LEAPS) in Pakistan and the Young Lives study in Ethiopia, India (states of Andhra Pradesh and Telangana), and Peru.⁷ Since the early 2000s, the two studies have regularly tracked children from early childhood into young adulthood, collecting repeated measures of household socioeconomic status, cognitive skills, educational attainment, and early-career labor market outcomes, among other data. This panel structure allows us to link FL to later-life outcomes for recent cohorts. We summarize key features of the sample and data here, with a detailed discussion in Appendix B.

Young Lives. We use data from the “older cohort” born in 1994/95, which has been surveyed in seven survey rounds between 2002 and 2023/24. We use the 2006 round, when children were about 12 years old, to obtain household-based childhood test scores and the 2023/24 survey round to measure long-term outcomes. By 2023/24, the overall attrition rate for this cohort was 24.3%, ranging from 16% in India to 30.4% in Ethiopia (Molina et al., 2025). Attrition is not predicted by age-12 test scores in any of the three countries (Appendix Table A1). We further restrict the Young Lives sample to the children for whom we have test scores at age 12.⁸ Our analytic sample includes 1,897 children in this cohort who were tested in 2006, have non-missing baseline socioeconomic controls, and were re-interviewed in 2023/24.

LEAPS. The LEAPS study collected panel data on education in 112 villages in three districts of Punjab province between 2003 and 2018/19 (Andrabi et al., 2007). Within sampled villages, the LEAPS team surveyed all primary schools within a feasible walking distance in 2003. They also independently sampled 1,807 households from the village household listing that

⁷The Young Lives study also collects data in Vietnam; however, data were not collected for this sample in the latest wave of data collection due to a change in governmental procedures for the international transfer of personal data. Since we focus on outcomes measured in the latest round, we exclude Vietnam from our analysis.

⁸About 95% of children interviewed in the 2006 survey round have non-missing test scores.

year, which forms the basis of our analytic sample. Our analytic LEAPS sample consists of 955 children in the household survey who have at least one test score between 2003 and 2006, non-missing baseline socioeconomic controls, and were re-observed in 2018/19.⁹ Most children in this sample were between ages 7 and 12 in 2003 (median age 10), and the child-weighted testing age is 11.2 years. Of those re-observed in 2018/19, 842 (88%) were interviewed in person. Where in-person interviews were not feasible, the team conducted phone surveys (1.7%) or obtained information through family members (10.2%). Similar to other countries, attrition is not predicted by childhood test scores (Appendix Table A1).

2.3 Key Measures

Test Scores. Young Lives and LEAPS both measure childhood learning using independent, low-stakes assessments administered at approximately ages 10–12. Within each study, repeated anchor items allow us to estimate item response theory (IRT) scores on a common latent scale across survey rounds. Because the two studies do not share linking items, however, test scores cannot be placed on a single cross-study scale. We therefore compare test-score gradients within each study, rather than interpreting score levels as directly comparable across LEAPS and Young Lives.

In LEAPS, test scores come from school-level tests in English, Urdu, and Mathematics between 2003 and 2006 as well as household-based tests in 2011 (Das et al., 2022). Using repeated anchor items, we estimate subject-specific IRT models and recover expected a posteriori (EAP) ability estimates on vertically linked scales across the 2003–2011 rounds. Our main childhood test-score measure averages the subject-specific IRT scores observed between 2003 and 2006, the period we treat as the childhood learning window. In the 2018/19 follow-up, the team administered cognitive tests to respondents who completed in-person surveys, when they were around age 24.5. The 2018 IRT scores are estimated on a separate adult scale. In the empirical analysis, we standardize childhood scores relative to the LEAPS analytic-sample distribution, and adult scores relative to the distribution among analytic-sample respondents with non-missing adult assessments.

In Young Lives, the team consistently administered home-based mathematics and language tests to both cohorts. For cross-country comparability, we focus on mathematics scores. We observe test scores for the older cohort from approximately ages 12 to 19 and for the younger cohort from approximately ages 8 to 15. As in LEAPS, scores are jointly calibrated across cohorts, countries, and rounds, so that EAP scores lie on a common scale (Espinoza Revollo and Scott, 2022). We measure FL using the older cohort’s scores at age 12 and adult skills using their scores at age 19, the last year for which we have test scores for this sample. We

⁹Half of our sample has test scores in all four years, and 94% was tested at least twice during this period.

standardize childhood scores relative to the Indian older cohort analytic-sample distribution and adult scores relative to the subset of the Indian analytic sample with non-missing adult scores.

Educational Attainment. We construct four measures of educational attainment: years of schooling, current enrollment, high school completion, and college/vocational attendance. In Pakistan, all four measures are constructed from the latest survey round. In Young Lives, current enrollment is measured in the latest round. For the other outcomes, we combine education histories collected in 2016 and the latest round, retaining the most complete or highest reported attainment. Years of schooling is based on the highest grade completed in formal education (including post-secondary education), or the current grade for respondents still enrolled. Current enrollment indicates whether the respondent was enrolled in formal education at the time of the survey; at these ages, this primarily reflects tertiary enrollment. High school completion is an indicator for completing upper-secondary schooling, and college/vocational attendance captures whether the respondent ever attended a technical, vocational, or university program.

Labor Market Outcomes. We construct two labor-market outcomes: work status and earnings. In LEAPS, work status indicates whether the respondent engaged in any paid activity in the past month. In Young Lives, work status indicates whether the respondent worked for at least one hour in the past week on a household farm, for someone outside the household, or in a self- or family-owned enterprise, regardless of whether they were paid. Earnings are based on reported monthly labor earnings, paid in cash or in-kind, from the respondent’s main activity in Young Lives and from the main activity plus agricultural earnings, if any, in LEAPS.¹⁰ We convert earnings to 2023 PPP US dollars to facilitate cross-country comparisons. In the regression analyses, we use log monthly earnings, restricting the sample to respondents with positive earnings and paying careful attention to any selection that this restriction may introduce. For the development accounting exercise (Section 4), we winsorize earnings among workers at the 1st and 99th percentiles within each country-cohort, separately by gender. Given respondents’ ages, these measures capture early-career labor market outcomes.

Socioeconomic Status. We construct three baseline socioeconomic measures: mother’s education, father’s education, and household wealth. Mother’s and father’s education are indicators for whether each parent has any formal education. Household wealth is measured using a country-specific asset index constructed from the first principal component of household durables and housing characteristics. We estimate this index separately for each country and then standardize it using each country’s analytic-sample distribution.¹¹

¹⁰In LEAPS, some low-intensity informal work reported by women (for example, home-based tutoring, teaching, or sewing) is not coded as “working” in the employment-status indicator, but any reported earnings from these activities are retained in the earnings measure.

¹¹In LEAPS, where multiple sampled children may share a household, we calculate the standardization moments across unique households represented by at least one analytic-sample child.

3 Main Results

3.1 Regression Results

We begin by examining the association between childhood test scores and later-life educational attainment and labor market outcomes. In each country, we estimate:

$$Y_i = \beta C_i + \alpha_a + \lambda_d + \omega_m + X_i' \gamma + \varepsilon_i \quad (1)$$

where Y_i is the later-life outcome of interest for individual i and C_i is the childhood test score for that individual. We include age fixed effects (α_a), cluster fixed effects (λ_d)¹², and gender fixed effects (ω_m). We also include mother’s education, father’s education, and household wealth as baseline socioeconomic controls (X_i). Finally, we cluster standard errors at the village level in Pakistan and at the community level in Ethiopia, India, and Peru.

Figure 1 examines the correlation between test scores at age 12 and educational attainment at ages 24.5 (Pakistan) and 28.5 (Ethiopia, India, and Peru). In all samples, FL is associated with substantial increases in educational attainment: A 1σ improvement in age-12 test scores is associated with 1.1–1.5 additional years of education across samples (Appendix Table A2), 9.3–13.5 percentage points higher probability of having attended some college/vocational education (Appendix Table A3), and a 7.6–14 percentage points higher probability of having finished higher secondary education (Appendix Table A4). At this age, most sampled individuals have completed formal education, and test scores are not correlated with enrollment (with the exception of Pakistan, where individuals are younger).

We investigate the correlation between FL and labor market outcomes in Table 1 for each country, for the full sample as well as by gender. Columns 1-4 focus on the probability of working. Test scores are not significantly associated with employment in the overall sample in any country, or for most gender $\times$ country subgroups. Young women are much less likely to report working in all countries than men of similar ages: about 80% of young men in Pakistan and Ethiopia, and over 90% of young men in India and Peru, report working, whereas the figures for women range from 7.5% in Pakistan to 71.4% in Peru.

Columns 5-8 of Table 1 focus on individuals who report positive earnings and present the core associations underpinning our analyses. Among respondents with positive earnings, a 1σ increase in FL corresponds to a 7.6–12.8 log-point increase (or approximately an 8–14% increase) in earnings in Pakistan, India, and Peru. Ethiopia stands out as an exception, with no statistically significant relationship between FL and earnings at age 28. It is plausible that this finding reflects the exceptional conditions in Ethiopia resulting from intense civil

¹²Equivalent to woredas in Ethiopia, mandals in India, and districts in Pakistan and Peru.

conflict in recent years. We do not find evidence of significant heterogeneity by gender in these correlations, although this mostly reflects the imprecise estimates for women.

This positive correlation between test scores and earnings could be driven by individuals at particular ranges of ability (e.g., high absolute test scores could be important primarily for predicting high-wage formal employment). We investigate this possibility in Figure 2, which presents a binned scatterplot with linear fits for all four countries, focusing on men and women with positive earnings. In India, Pakistan, and Peru, the relationship between log earnings and test scores is well described by a linear fit for men (left panel). Although our samples are not sufficiently powered to test for nonlinearities, we do not see evidence suggestive of a positive earnings-test scores gradient only after particular thresholds. The relationship remains substantially positive in these three countries after conditioning on household SES and parental education. There are substantial intercept differences between Peru and India on the one hand, and Pakistan and Ethiopia on the other. For women (right panel) with positive earnings, we observe a positive correlation between log earnings and FL in all countries (although the slope is not significant in any country, reflecting smaller sample sizes). Intercept differences between countries are substantially larger for women than for men.

3.2 Explaining the Skills-Earnings Gradient

We next provide a predictive decomposition to assess if adult test scores and attainment, along with their downstream association with earnings, can explain the entire association between FL and adult earnings. If they span most of the predictive gradient, they may be useful candidate surrogates, although their validity for causal extrapolation would need to be established in experimental settings (Athey et al., 2025). We focus on years of schooling and adult test scores since they have been the main focus of prior research (Ozawa et al., 2022).

Let E_i denote adult log earnings, C_i childhood test scores, A_i adult test scores, and S_i schooling. One way to assess whether the predictive relationship between childhood test scores and adult earnings can be accounted for by the observed persistence of childhood achievement into adult test scores and schooling is to regress adult earnings on adult test scores, schooling and child test scores. However, adult test scores and schooling are downstream outcomes of childhood achievement, and test scores (both child and adult) are measured with error. This makes it difficult to interpret the coefficients given unknown biases. Instead, we adopt a predictive accounting approach based on separately estimated bivariate relationships and use estimates from the following regressions:

$$E_i = \alpha_C + \pi C_i + u_i, \quad (2)$$

$$E_i = \alpha_A + \beta A_i + v_i, \quad (3)$$

$$E_i = \alpha_S + \theta S_i + w_i, \quad (4)$$

$$A_i = \mu_A + \gamma C_i + r_i, \quad (5)$$

$$S_i = \mu_S + \delta C_i + q_i. \quad (6)$$

where E_i is log earnings for individual i , C_i is the child test scores, A_i is the adult test score, and S_i is years of schooling.

We also estimate the correlation between adult test scores and schooling:

$$\rho = \text{Corr}(A_i, S_i). \quad (7)$$

A naive chain-rule accounting exercise would attribute $\pi^{\text{naive}} = \beta\gamma + \theta\delta$ of the childhood-score earnings gradient to the two observed pathways but this expression generally overstates the contribution of the observed pathways when adult test scores and schooling are positively correlated, since the two terms may reflect some of the same underlying child-score variation. To account for this overlap, we use the covariance-adjusted explained component:

$$\pi^{\text{explained}} = \frac{\beta\gamma + \theta\delta - \rho(\theta\gamma + \beta\delta)}{1 - \rho^2}. \quad (8)$$

Equation (8) is the predictive gradient in earnings with respect to childhood test scores implied by the linear projection of earnings on the two observed adult human-capital measures, while allowing those two measures to be correlated.¹³

¹³Consider the linear projection

$$E_i = b_A A_i + b_S S_i + \varepsilon_i.$$

With standardized variables, the projection coefficients are

$$b_A = \frac{\beta - \rho\theta}{1 - \rho^2}, \quad b_S = \frac{\theta - \rho\beta}{1 - \rho^2},$$

where $\beta = \text{Corr}(E, A)$, $\theta = \text{Corr}(E, S)$, and $\rho = \text{Corr}(A, S)$. Since $\gamma = \text{Corr}(A, C)$ and $\delta = \text{Corr}(S, C)$, the childhood-score gradient in earnings implied by these two adult pathways is

$$b_A \gamma + b_S \delta = \frac{(\beta - \rho\theta)\gamma + (\theta - \rho\beta)\delta}{1 - \rho^2} = \frac{\beta\gamma + \theta\delta - \rho(\theta\gamma + \beta\delta)}{1 - \rho^2}.$$

Thus, the covariance adjustment subtracts the component that would otherwise be double-counted because adult test scores and schooling are correlated.

We then compare the total reduced-form childhood-score earnings gradient π to the covariance-adjusted explained component $\pi^{\text{explained}}$. The residual component is

$$\pi^{\text{res}} = \pi - \pi^{\text{explained}} = \pi - \frac{\beta\gamma + \theta\delta - \rho(\theta\gamma + \beta\delta)}{1 - \rho^2}. \quad (9)$$

A positive residual, π^{res} , indicates that childhood scores contain predictive information about earnings that is not spanned by the linear projections on measured schooling and later test scores. This residual need not represent a separate causal pathway: it may also reflect measurement error, omitted common determinants, functional-form restrictions, or selection into the earnings sample.

We implement this decomposition for men in the four countries.

Table 2 reports two specifications. Within each country and panel, we first residualize all four variables with respect to the relevant controls and then standardize the residuals. Panel A uses age and cluster fixed effects, while Panel B additionally includes baseline household wealth and parental education. Both panels use the same country-specific estimation sample, so that differences between panels reflect the inclusion of SES controls rather than changes in sample composition. We provide percentile confidence intervals from 1,000 cluster-bootstrap replications, where we resample villages in Pakistan and communities in Young Lives.

Prior to the composite decomposition in Table 2, Appendix Table A5 reports the component estimates corresponding to Equations (2)–(7). As expected, childhood test scores are strongly predictive of adult test scores in all four samples: with SES controls, a one-standard-deviation increase in childhood scores is associated with 0.32 to 0.55 standard deviations higher adult scores.¹⁴ Childhood scores also predict years of schooling, with coefficients ranging from 0.19 in Peru to 0.45 in Pakistan. Adult test scores and schooling are positively correlated in every country, with correlations ranging from 0.34 to 0.68 after controlling for SES, so a naive sum of the two pathways would double-count part of the same underlying variation. The adult-score–earnings gradient is positive in Pakistan, India, and Peru, but close to zero in Ethiopia, mirroring the child-score earnings gradient. By contrast, the schooling–earnings gradient is positive in all four countries.

We then combine these components in Table 2. The results are informative in two ways. First, the total child-score earnings gradient is positive in Pakistan, India, and Peru, but close to zero in Ethiopia. With baseline SES controls, the standardized gradient ($\hat{\pi}$) is 0.098 in Pakistan,

¹⁴For comparison, these inter-period persistence coefficients are of a similar magnitude as those reported for test scores in successive years. See, for example, [Andrabi et al. \(2011\)](#), using these data in Pakistan, or [Muralidharan and Singh \(2020, 2025\)](#) in different states in India. The test score persistence we observe here appears meaningfully larger than might be expected from geometric decay over even moderate time periods of 2–3 years. This suggests that test scores measured at the end of primary schooling may be particularly important in consolidating learning; see also the analysis of learning trajectories in [Bau et al. \(2021\)](#).

0.134 in India, and 0.203 in Peru, compared with 0.009 in Ethiopia. The Ethiopian estimates are therefore difficult to interpret as a decomposition of the child-score earnings gradient, since there is little to decompose. Second, interpreted descriptively, the observed adult-test-score and schooling pathways (with SES controls) predicted component ($\hat{\pi}^{\text{expl}}$) is 0.049 in Pakistan, 0.076 in India, and 0.037 in Peru. These estimates imply that the two observed pathways account for roughly 50% of the total gradient in Pakistan, 57% in India, and 18% in Peru. The remaining component is largest in Peru, where the residual estimate’s confidence interval excludes zero; in Pakistan and India, the residual components are positive but imprecisely estimated. Overall, the decomposition suggests that schooling and adult cognitive skills capture an important part of the association between childhood learning and adult earnings, but do not fully account for it in the countries where the total earnings gradient is positive. The predictive decomposition exercise suggests that the link between FL and earnings may be *stronger* than the association between earnings and years of schooling in India, Pakistan, and Peru, even after including SES controls.

4 Development Accounting

We now investigate the macroeconomic implications of our empirical results for our understanding of the overall contribution of FL and schooling to cross-country earnings differences.

4.1 Theoretical Framework

We consider a simple growth accounting framework in which human capital determines incomes through both a quantity (years of schooling) and a quality (test scores) components (e.g., [Hanushek et al. \(2017\)](#)). Output per capita is a function of capital, labor, and total factor productivity:

$$Y_c = K_c^\alpha (A_c L_c)^{1-\alpha} \quad (10)$$

where c indexes countries. Labor is a function of the quantity and quality of schooling:

$$L_c = e^{rS_c + \phi Q_c} \quad (11)$$

where S_c is average years of schooling, r is the Mincerian return to a year of schooling, Q_c is the average quality of schooling measured in standard deviations of test scores, and ϕ is the return to a standard deviation of test scores.

Taking logs and differencing, the log gap in output per capita between two countries can be decomposed into the contributions of the physical-capital-to-output ratio, total factor

productivity, and human capital:

$$1 = \frac{\alpha}{1 - \alpha} \frac{\Delta \log \frac{K}{Y}}{\Delta \log Y} + \frac{\Delta \log A}{\Delta \log Y} + \frac{r\Delta S + \phi\Delta Q}{\Delta \log Y} \quad (12)$$

as is standard in the literature (Hsieh and Klenow, 2010). The share of output per capita differences explained by schooling and learning is thus:

$$\text{Share}_L = \underbrace{\frac{r\Delta S}{\Delta \log Y}}_{\text{Quantity of Schooling}} + \underbrace{\frac{\phi\Delta Q}{\Delta \log Y}}_{\text{Quality of Schooling}} \quad (13)$$

4.2 Empirical Application

We bring this framework to the data by focusing on the difference between Ethiopia and India (Andhra Pradesh and Telangana), which display the largest earnings gap in our sample. Average earnings differ by a factor of almost 2.5 between the two countries: they are about \$500 in Ethiopia compared to \$1,200 in India. Indian cohorts display both greater years of schooling (12.2 years vs. 10.6 years) and greater standardized test scores (0.64 vs. 0).

We consider three different decompositions. We first investigate the joint contribution of the quantity and quality of schooling. For this specification, we use conditional estimates of the returns to schooling and learning, that is, we focus on returns to schooling obtained after controlling for learning and returns to learning obtained after controlling for years of schooling. We then consider specifications with either schooling alone or learning alone, using raw returns that do not control for the other dimension of human capital in the regression. These alternative specifications are particularly useful to assess the magnitude of policies that would specifically close the learning or schooling gap, accounting for the fact that improving learning can increase years of schooling and vice versa.

For each decomposition, we consider three alternative specifications corresponding to more or less optimistic scenarios on the returns to schooling and learning (see Table 3, Panel A). In our benchmark specification, we assume conditional returns to schooling and learning of 4% and 8%, respectively, and unconditional returns of 5% and 10%. These returns fall in the middle of returns estimated in our four countries. In the conservative scenario, we assume much lower returns of 2-7%, closer to those observed in Ethiopia and considerably lower than those typically used in the literature. Finally, we consider larger returns in our optimistic scenario, 7% for schooling and 10% for learning in the conditional specification.

4.3 Results

Table 3, Panel B shows the results of the development accounting decomposition. In our benchmark specification, learning and schooling combined can account for approximately 14% of the wage gap between young Ethiopians and Andhraites. Schooling accounts for 8 percentage points of this contribution (or about 55% of the total), while learning accounts for 6 points (or about 45% of the total). The total contribution of human capital ranges from 8% in the conservative scenario to 22% in the optimistic scenario. The last two rows show the results for schooling and learning unconditionally. Schooling alone can account for 6-16% of the wage gap depending on the specification. FL alone can account for 5-12%.

These numbers are smaller than those found in the literature, mainly due to the low returns to schooling and learning that we use (e.g., [Angrist et al. \(2021\)](#); [Schoellman \(2012\)](#)). These lower returns are likely related to the young sample that forms the basis of our analysis. For instance, [Deming \(2025\)](#) documents that the returns to schooling and cognitive skills increase by a factor of about 2.3 and 2.7, respectively, between ages 26 and 60 (see also [Hanushek et al. \(2015\)](#)). After adjusting our returns by these factors, schooling and learning can account for as much as 50% of the earnings gap, closer to the orders of magnitude found in the literature and even slighter larger. Learning alone can account for about 30%.

These results point to three tentative conclusions. First, human capital can account for a non-negligible fraction of the large income gap observed between Ethiopia and India. Second, this contribution is likely underestimated, given the young sample that forms the basis of our analysis, the significant noise associated with the measurement of schooling and learning, and potential positive human capital externalities. Third, both schooling and learning matter: Ethiopia and India differ on both dimensions of human capital, and we estimate positive returns to each of them even after conditioning on one another.

5 Conclusion

The “global learning crisis” is widely recognized as a major challenge for education systems in many low- and middle-income countries today ([World Bank, 2018](#)). Yet, despite systematic efforts among some countries and some funding bodies ¹⁵ as well as substantial research identifying effective interventions to improve FL ([Angrist et al., 2025](#)), there has not been a clear commitment to improving FL across LMICs. Part of the reason for this may be the lack of direct evidence on these returns in the absence of panel data linking childhood test scores to adult earnings.

If this has indeed been the stumbling block, our results can help advance the debate with three key insights. First, our results reassure us that, at least at the level of associations, FL

¹⁵India’s National Education Policy recognizes universal FL as a key goal, as does the Gates Foundation. See ([Muralidharan and Singh, 2021](#)) and ([Bill & Melinda Gates Foundation, 2026](#)).

skills and earnings are correlated across the multiple settings that we examine (barring the anomalous example of Ethiopia at the time of conflict). The FL-earnings gradient emerges by the time respondents reach their late-20s and may grow larger if earnings for the highly educated follow a similar trajectory as that in OECD economies (Bhuller et al., 2017). Reassuringly, Andrabi et al. (2026) have computed causal returns in Pakistan by following students who were part of an experiment in primary school, and find even *higher* estimates. Given that the gradient in Pakistan is much *lower* than what we find in India and Peru, our associations can be regarded as indicative of the size of these returns, much as Mincerian returns provide a benchmark for a substantial literature providing causal estimates of the return to education.

Second, returns to FL may be larger than what can be explained by the effect of FL on schooling and skills in adulthood. Given the long lag between the measurement of FL skills and the realized returns, understanding what explains the FL-earnings gradient will be important for future research. This may require (i) the validated measurement of a broader set of skills in LMIC settings (e.g. Laajaj and Macours (2021) and Danon et al. (2024)); (ii) the returns to such skills in the labor market in adulthood, and (iii) the process by which FL translates into future skills. Such a program could help develop a full path of surrogate outcomes to estimate the lifetime effects of FL interventions (Athey et al., 2025). Looking solely at years of schooling and narrow measures of cognitive skills in adulthood appears unlikely to be adequate.

Third, recognizing that policy attention to FL also comes from potential effects on macroeconomic growth, we show that for a benchmark decomposition, the figures we get are in line with the broader development accounting literature. The economic case for focusing on skills in primary schooling remains strong even when focusing on actual comparable measures of childhood skills and their returns rather than, e.g., indirect measures backed out from combining harmonized international test scores with returns to cognitive skills in adulthood.

Our descriptive evidence in this paper helps to empirically ground discussions on potential returns from improving skills in primary schooling. In the absence of reliable administrative data in LMICs, long-running panel studies remain our only source for direct benchmarks to quantify how much a better teacher or a better school could be worth in net present value. Developing and maintaining such panels should be a high priority not only as a basic research infrastructure but, also, for the applied purpose of providing core benchmarks to substantial resource allocations.

References

- Andrabi, T., Bau, N., Das, J., Ijaz Khwaja, A., and Michaud-Leclerc, C. (2026). The long-term effects of school quality in a low-income country: Evidence from 15 years of data. Technical report, Georgetown University.
- Andrabi, T., Das, J., Ijaz Khwaja, A., and Zajonc, T. (2011). Do value-added estimates add value? accounting for learning dynamics. *American Economic Journal: Applied Economics*, 3(3):29–54.
- Andrabi, T., Das, J., Khwaja, A., Vishwanath, T., and Zajonc, T. (2002). Test feasibility survey pakistan: Education sector. *Cambridge, MA: Harvard Kennedy School Working Paper*.
- Andrabi, T., Das, J., and Khwaja, A. I. (2022). The learning and educational achievement in pakistan schools (leaps) longitudinal dataset, 2004-2011. Released November 1st 2022. Version 11-2022.
- Andrabi, T., Das, J., Khwaja, A. I., Vishwanath, T., and Zajonc, T. (2007). Pakistan: Learning and educational achievements in punjab schools (leaps): Insights to inform the education policy debate. Technical report, World Bank, Washington, DC.
- Angrist, N., Djankov, S., Goldberg, P. K., and Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, 592(1):403–408.
- Angrist, N., Evans, D. K., Filmer, D., Glennerster, R., Rogers, F. H., and Sabarwal, S. (2025). How to improve education outcomes most efficiently? a review of the evidence using a unified metric. *Journal of Development Economics*, 172:103192.
- Athey, S., Chetty, R., Imbens, G. W., and Kang, H. (2025). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. *Review of Economic Studies*, 93(4):2284–2312.
- Bau, N., Das, J., and Chang, A. Y. (2021). New evidence on learning trajectories in a low-income setting. *International Journal of Educational Development*, 84:102430.
- Beegle, K., De Weerd, J., and Dercon, S. (2011). Migration and economic mobility in tanzania: Evidence from a tracking survey. *Review of Economics and Statistics*, 93(3):1010–1033.
- Bhuller, M., Mogstad, M., and Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4):993–1030.

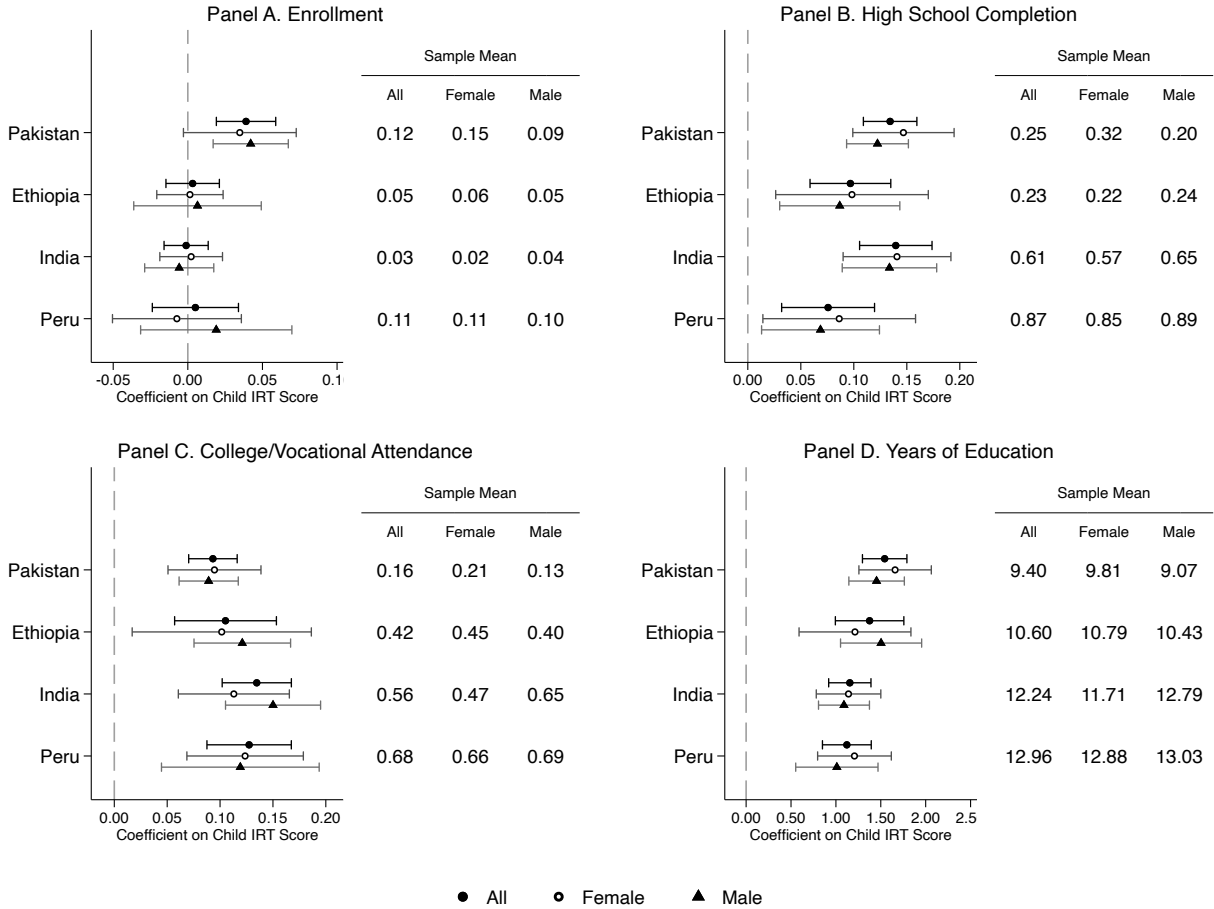
- Bill & Melinda Gates Foundation (2026). Global education program. <https://www.gatesfoundation.org/our-work/programs/global-growth-and-opportunity/global-education-program>. Accessed May 13, 2026.
- Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of economic growth*, 1:679–741.
- Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679.
- Chua, K. (2017). Skill achievement and returns in developing countries: Evidence from adult skills surveys. *European Journal of Education*, 52(4):498–510.
- Crawford, L. (2026). The economic returns to foundational literacy and numeracy: Evidence from indonesia. *The Journal of Development Studies*.
- Danon, A., Das, J., de Barros, A., and Filmer, D. (2024). Cognitive and socioemotional skills in low-income countries: Measurement and associations with schooling and earnings. *Journal of Development Economics*, 168:103132.
- Das, J., Singh, A., and Chang, A. Y. (2022). Test scores and educational opportunities: Panel evidence from five low-and middle-income countries. *Journal of Public Economics*, 206:104570.
- Deming, D. (2025). Why do wages grow faster for educated workers? *Journal of Labor Economics*. Just Accepted.
- Escobal, J. and Flores, E. (2008). An assessment of the young lives sampling approach in peru. Technical Note 3, Young Lives, Oxford.
- Espinoza Revollo, P. and Scott, D. (2022). Cognitive and achievement tests in the young lives study. Technical Note 53, Young Lives, Oxford, UK.
- Fredriksson, P., Öckert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly journal of economics*, 128(1):249–285.
- Glewwe, P., Song, Y., and Zou, X. (2022). Labor market outcomes, cognitive skills, and noncognitive skills in rural china. *Journal of Economic Behavior & Organization*, 193:294–311.
- Hanushek, E., Ruhose, J., and Woessman, L. (2017). Knowledge capital and aggregate income differences: Development accounting for us states. *American Economic Journal: Macroeconomics*, 9(4):184–224.

- Hanushek, E. A., Schwerdt, G., Wiederhold, S., Woessmann, L., et al. (2015). Returns to skills around the world. *European Economic Review*, 73:103–130.
- Hanushek, E. A. and Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of economic literature*, 46(3):607–668.
- Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly journal of economics*, 135(3):1209–1318.
- Hendricks, L. and Schoellman, T. (2018). Human capital and development accounting: New evidence from wage gains at migration. *The Quarterly Journal of Economics*, 133(2):665–700.
- Hsieh, C.-T. and Klenow, P. (2010). Development accounting. *American Economic Journal: Macroeconomics*, 2(1):207–223.
- Kumra, N. (2008). An assessment of the young lives sampling approach in andhra pradesh, india. Technical Note 2, Young Lives, Oxford.
- Laajaj, R. and Macours, K. (2021). Measuring skills in developing countries. *Journal of Human Resources*, 56(4):1254–1295.
- Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2):407–437.
- Manuelli, R. E. and Seshadri, A. (2014). Human capital and the wealth of nations. *American economic review*, 104(9):2736–2762.
- McGowan, C. R., Cassidy-Seyoum, S. A., Ekoriko, P., Alhaffar, M., Cassini, L., Palmer, J., Warsame, A., and Checchi, F. (2025). Conflict-attributable mortality in tigray region, ethiopia: Evidence from a survey of the tigrayan diaspora. *Population Health Metrics*, 23(1):19.
- Molina, M., Favara, M., Sánchez, A., and Woodman Deza, A. (2025). Young lives attrition report: Round 7. Technical Note 58, Young Lives, Oxford.
- Muralidharan, K. and Singh, A. (2020). Improving public sector management at scale? experimental evidence on school governance india. Technical report, National Bureau of Economic Research.
- Muralidharan, K. and Singh, A. (2021). India’s new national education policy: Evidence and challenges. *Science*, 372(6537):36–38.
- Muralidharan, K. and Singh, A. (2025). Adapting for scale: Experimental evidence on technology-aided instruction in india. Technical report, National Bureau of Economic Research.

- Outes-Leon, I. and Sanchez, A. (2008). An assessment of the young lives sampling approach in ethiopia. Technical Note 1, Young Lives, Oxford.
- Ozawa, S., Laing, S. K., Higgins, C. R., Yemeke, T. T., Park, C. C., Carlson, R., Ko, Y. E., Guterman, L. B., and Omer, S. B. (2022). Educational and economic returns to cognitive ability in low-and middle-income countries: A systematic review. *World development*, 149:105668.
- Schoellman, T. (2012). Education quality and development accounting. *The Review of Economic Studies*, 79(1):388–417.
- Valerio, A., Sanchez Puerta, M. L., Tognatta, N., and Monroy-Taborda, S. (2016). Are there skills payoffs in low- and middle-income countries? empirical evidence using step data. Policy Research Working Paper 7879, World Bank.
- World Bank (2018). *World Development Report 2018: Learning to Realize Education’s Promise*. World Bank, Washington, DC.
- World Bank, UNESCO, UNICEF, FCDO, USAID, and Bill & Melinda Gates Foundation (2022). The state of global learning poverty: 2022 update. Technical report, World Bank, UNESCO, UNICEF, FCDO, USAID, and Bill & Melinda Gates Foundation. Conference edition, June 23, 2022.

Main Exhibits

Figure 1: Correlation with Educational Attainment



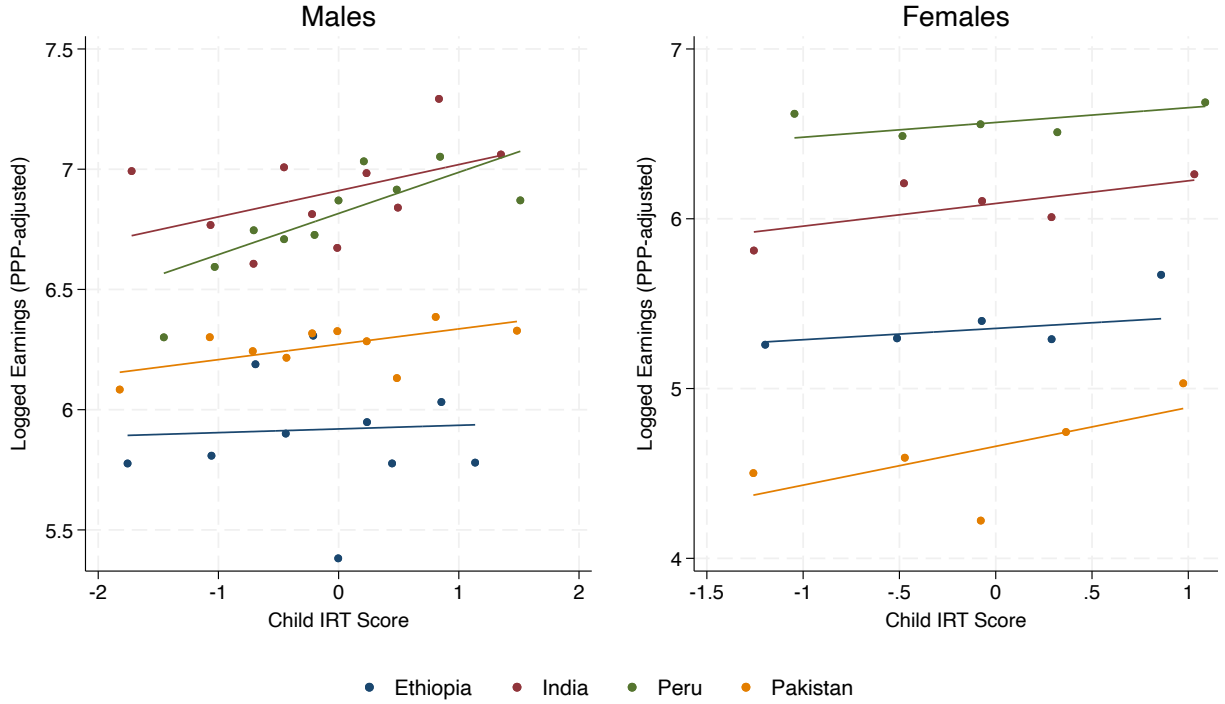
Notes: This figure reports the association between childhood test scores and adult educational attainment in the latest available survey round. Panels A–D report estimates for enrollment at the time of the survey, high school completion, ever attending college or vocational education, and years of schooling, respectively. The plotted points are coefficients on childhood test scores, estimated separately for the full sample, females, and males. Horizontal bars report 95 percent confidence intervals, and the right-hand table in each panel reports the corresponding outcome means for the full sample, females, and males. Sample means and coefficients in Panels A–C are reported as proportions. The sample includes respondents from LEAPS in Pakistan and the Young Lives older cohort in Ethiopia, India, and Peru. In LEAPS, childhood test scores are the average of mathematics, Urdu, and English IRT scores from tests administered between 2003 and 2006, and the child-weighted mean testing age is 11.2 years. In Young Lives, childhood test scores are mathematics IRT scores from 2006, when older-cohort respondents were around age 12. LEAPS scores are standardized using the LEAPS analytic-sample distribution, while Young Lives scores are standardized using the Indian older-cohort analytic-sample distribution. Because LEAPS and Young Lives do not share linking items, test-score levels are not directly comparable between the two studies. All specifications control for baseline household wealth, measured using a within-country standardized asset index; parental education, measured as indicators for whether each parent has any formal education; age fixed effects; and district fixed effects for LEAPS or sampling-cluster fixed effects for Young Lives. Full-sample specifications additionally include gender fixed effects. Standard errors are clustered at the village level for LEAPS and at the community level for Young Lives.

Table 1: Correlation between Labor Market Outcomes and Child Test Scores

	Working				Logged Earnings			
	(1) Pakistan	(2) Ethiopia	(3) India	(4) Peru	(5) Pakistan	(6) Ethiopia	(7) India	(8) Peru
Panel A: Full Sample								
Child IRT Score	-0.008 (0.011)	0.028 (0.024)	-0.021 (0.016)	0.027 (0.021)	0.076** (0.033)	0.047 (0.046)	0.126*** (0.044)	0.128*** (0.044)
Sample Mean	0.479	0.607	0.730	0.829	6.058	5.719	6.615	6.724
Adjusted R ²	0.537	0.227	0.187	0.106	0.313	0.125	0.277	0.126
Observations	955	610	801	486	471	300	503	384
Clusters	108	24	99	59	102	24	99	54
Panel B: Females Only								
Child IRT Score	0.030*** (0.011)	0.054 (0.034)	0.006 (0.025)	0.045 (0.045)	0.229 (0.138)	0.067 (0.120)	0.134 (0.083)	0.087 (0.099)
Sample Mean	0.075	0.393	0.545	0.714	4.736	5.322	6.075	6.571
Adjusted R ²	0.025	0.109	0.031	0.081	0.302	0.268	0.158	0.047
Observations	428	290	407	231	64	98	178	149
Clusters	106	24	97	52	47	21	78	40
Panel C: Males Only								
Child IRT Score	-0.044** (0.018)	0.001 (0.026)	-0.040** (0.016)	0.015 (0.017)	0.064* (0.033)	0.015 (0.068)	0.109** (0.051)	0.171*** (0.056)
Sample Mean	0.806	0.800	0.921	0.933	6.266	5.912	6.910	6.821
Adjusted R ²	0.059	0.111	0.065	0.017	0.024	-0.032	0.162	0.080
Observations	527	320	394	255	407	202	325	235
Clusters	105	22	94	48	100	22	94	46

Notes: This table reports the association between childhood test scores and labor market outcomes in the latest survey round. We report associations with the respondent's work status in columns (1)–(4) and log monthly earnings for respondents with positive earnings in columns (5)–(8). In column (1), for Pakistan, the dependent variable is an indicator for whether the respondent was engaged in any paid activity in the past month. For the Young Lives countries in columns (2)–(4), the dependent variable is an indicator for whether the respondent engaged in any paid or unpaid activity (including work on a household farm, work for someone outside the household, or work in a self- or family-owned enterprise) for at least one hour in the past week. In columns (5)–(8), earnings are converted to 2023 PPP dollars before being logged. In LEAPS, childhood test scores are the average of mathematics, Urdu, and English IRT scores from tests administered between 2003 and 2006, and the child-weighted mean testing age is 11.2 years. In Young Lives, childhood test scores are mathematics IRT scores from 2006, when older-cohort respondents were around age 12. LEAPS scores are standardized using the LEAPS analytic-sample distribution, while Young Lives scores are standardized using the Indian older-cohort analytic-sample distribution. Because LEAPS and Young Lives do not share linking items, test-score levels are not directly comparable between the two studies. All specifications control for baseline household wealth, measured using a within-country standardized asset index; parental education, measured as indicators for whether each parent has any formal education; age fixed effects; and district fixed effects for LEAPS or sampling-cluster fixed effects for Young Lives. Full-sample specifications additionally include gender fixed effects. Standard errors are clustered at the village level for LEAPS and at the community level for Young Lives. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2: Correlation between Logged Earnings and Child Test Scores



Notes: This figure reports binned scatterplots of log monthly earnings against childhood test scores separately for males (left panel) and females (right panel) with positive earnings in Pakistan and the Young Lives older cohort in Ethiopia, India, and Peru. Earnings are converted to 2023 PPP dollars before being logged. Plotted points are country-specific binned means using ten quantile-spaced bins for males and five quantile-spaced bins for females (reflecting the smaller female earnings sample), and fitted lines are estimated separately by country. Both figures control for age fixed effects, district fixed effects in LEAPS or sampling-cluster fixed effects in Young Lives, and baseline SES controls. Baseline SES controls include baseline household wealth, measured using a within-country standardized asset index, and parental education, measured as indicators for whether each parent has any formal education. In LEAPS, childhood test scores are the average of mathematics, Urdu, and English IRT scores from tests administered between 2003 and 2006, and the child-weighted mean testing age is 11.2 years. In Young Lives, childhood test scores are mathematics IRT scores from 2006, when older-cohort respondents were around age 12. LEAPS scores are standardized using the LEAPS analytic-sample distribution, while Young Lives scores are standardized using the Indian older-cohort analytic-sample distribution. Because LEAPS and Young Lives do not share linking items, test-score levels are not directly comparable between the two studies.

Table 2: Predictive Decomposition (Men Only)

	LEAPS (24.5y)	Young Lives (28.5y)		
	(1) Pakistan	(2) Ethiopia	(3) India	(4) Peru
Panel A: No SES controls				
Total gradient, $\hat{\pi}$	0.121	0.019	0.220	0.216
95% CI	[0.023, 0.216]	[-0.096, 0.122]	[0.098, 0.332]	[0.034, 0.394]
Predicted component, $\hat{\pi}^{\text{expl}}$	0.079	0.032	0.148	0.051
95% CI	[0.022, 0.138]	[-0.032, 0.086]	[0.083, 0.199]	[-0.018, 0.145]
Residual component, $\hat{\pi}^{\text{res}}$	0.042	-0.013	0.071	0.164
95% CI	[-0.033, 0.122]	[-0.112, 0.085]	[-0.026, 0.169]	[0.029, 0.274]
Observations	335	198	323	224
Panel B: With SES controls				
Total gradient, $\hat{\pi}$	0.098	0.009	0.134	0.203
95% CI	[0.005, 0.202]	[-0.105, 0.120]	[-0.001, 0.257]	[0.053, 0.375]
Predicted component, $\hat{\pi}^{\text{expl}}$	0.049	0.030	0.076	0.037
95% CI	[-0.006, 0.106]	[-0.032, 0.079]	[0.026, 0.132]	[-0.022, 0.126]
Residual component, $\hat{\pi}^{\text{res}}$	0.049	-0.021	0.059	0.166
95% CI	[-0.029, 0.130]	[-0.140, 0.092]	[-0.046, 0.161]	[0.043, 0.284]
Observations	335	198	323	224

Notes: This table reports a descriptive predictive decomposition of the association between childhood test scores and log monthly earnings for men with positive earnings. Columns (1)–(4) report estimates for Pakistan and the Young Lives older cohort in Ethiopia, India, and Peru, respectively. Earnings are converted to 2023 PPP dollars before being logged. Within each country and panel, log earnings, childhood test scores, adult test scores, and years of schooling are residualized and standardized before the decomposition. The total gradient, $\hat{\pi}$, is the standardized association between childhood test scores and log earnings. The predicted component, $\hat{\pi}^{\text{expl}}$, is the component predicted by the adult test-score and years of schooling pathways, allowing adult test scores and schooling to be correlated. The residual component is $\hat{\pi}^{\text{res}} = \hat{\pi} - \hat{\pi}^{\text{expl}}$. Panel A includes age fixed effects and district fixed effects for LEAPS or sampling-cluster fixed effects for Young Lives. Panel B additionally controls for baseline household wealth, measured using a within-country standardized asset index, and parental education, measured as indicators for whether each parent has any formal education. Both panels use the same country-specific sample with non-missing log earnings, childhood test scores, adult test scores, years of schooling, baseline SES controls, age, and district or sampling-cluster identifiers. Confidence intervals are constructed by repeating the decomposition in 1,000 cluster-bootstrap samples, resampling villages in LEAPS and communities in Young Lives, and reporting the 2.5th and 97.5th percentiles of the bootstrapped estimates. Because all variables entering the decomposition are standardized within country and panel, the reported gradients and components are in standard-deviation units.

Table 3: Development Accounting Results

	Conservative	Benchmark	Optimistic
Panel A. Returns			
Return to schooling conditional	2%	4%	7%
Return to learning conditional	5%	8%	10%
Return to schooling unconditional	3%	5%	8%
Return to learning unconditional	7%	10%	16%
Panel B. Human capital contribution			
Learning and schooling combined	7.8%	14.1%	21.6%
<i>of which: schooling</i>	4.0%	7.9%	13.9%
<i>of which: learning</i>	3.9%	6.2%	7.7%
Schooling alone	5.9%	9.9%	15.9%
Learning alone	5.4%	7.7%	12.4%

Notes: This table reports the results of a development accounting decomposition isolating the contributions of schooling and learning to the real earnings gap between Ethiopia and India. Panel A reports more or less conservative estimates of the returns to schooling and learning. Panel B reports the corresponding share of the earnings gap between Ethiopia and India that can be accounted for by schooling and learning with these returns. Return to schooling/learning conditional: return to a year of schooling/to a standard deviation of test score, estimated when including both schooling and test scores in the regression. Returns to schooling/learning unconditional: return to a year of schooling/to a standard deviation of test score, estimated without controlling for the other dimension of human capital in the regression. Learning and schooling combined: total contribution of learning and schooling, estimated using conditional returns to both. Schooling alone/Learning alone: contributions of learning and schooling alone, estimated using unconditional returns. Under our benchmark specification, learning and schooling combined can account for about 9% of the earnings gap between Ethiopia and India, of which 6-7 points are due to schooling and 2-3 points are due to learning. Schooling alone can rationalize 8% of the earnings gap, while foundational learning alone can rationalize 3%.

Appendix A: Tables and Figures

Table A1: Likelihood of Being Observed at Endline

	Pakistan		Ethiopia		India		Peru	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Bivariate	Multiple	Bivariate	Multiple	Bivariate	Multiple	Bivariate	Multiple
Child IRT Score	0.004 (0.014)	0.003 (0.014)	0.027 (0.021)	0.032 (0.021)	0.011 (0.010)	0.015 (0.010)	-0.024 (0.021)	-0.021 (0.020)
Mother has Some Education	0.018 (0.025)	0.019 (0.029)	-0.000 (0.021)	-0.004 (0.025)	-0.006 (0.031)	0.005 (0.035)	0.011 (0.083)	0.011 (0.077)
Father has Some Education	0.007 (0.024)	0.008 (0.027)	-0.011 (0.025)	-0.009 (0.027)	-0.018 (0.021)	-0.013 (0.026)	0.218 (0.133)	0.232* (0.135)
Baseline Asset Index	-0.009 (0.015)	-0.013 (0.015)	-0.028 (0.018)	-0.034** (0.016)	-0.016 (0.014)	-0.019 (0.016)	-0.039 (0.038)	-0.042 (0.038)
Male	0.005 (0.021)	0.005 (0.021)	0.038 (0.039)	0.033 (0.038)	0.008 (0.029)	0.005 (0.029)	-0.044 (0.029)	-0.043 (0.028)
Urban							-0.016 (0.060)	0.027 (0.065)
Age at Baseline	-0.005 (0.006)	-0.005 (0.007)	0.009 (0.037)	0.007 (0.038)	-0.020 (0.028)	-0.018 (0.028)	-0.048 (0.048)	-0.048 (0.046)
Sample Mean	0.853	0.853	0.718	0.718	0.857	0.857	0.756	0.756
Cluster/District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
P-value of F-test	.	0.894	.	0.143	.	0.670	.	0.287
Adjusted R ²	.	0.020	.	0.107	.	0.067	.	0.010
Observations	.	1119	.	850	.	935	.	643

Notes: This table tests whether baseline characteristics predict observation in the long-term follow-up surveys. The dependent variable equals one if the respondent is observed at endline and zero otherwise. Endline refers to the 2018/19 LEAPS survey in Pakistan and the 2023/24 Young Lives survey in Ethiopia, India, and Peru. The Young Lives estimates are restricted to the older cohort. Columns (1)–(2), (3)–(4), (5)–(6), and (7)–(8) report results for Pakistan, Ethiopia, India, and Peru, respectively. Odd-numbered columns report separate bivariate regressions of endline observation on each listed characteristic, while even-numbered columns include the characteristics jointly. Within each country, the bivariate and multiple specifications use the same analytic sample, restricted to respondents with non-missing values for all baseline characteristics. The predictors are a standardized childhood IRT score, indicators for whether the mother and father have any formal education, a within-country standardized household asset index, gender, urban residence, and age at baseline. In LEAPS, childhood test scores are based on tests administered between 2003 and 2006, while the remaining baseline characteristics are measured in 2003. In the Young Lives older cohort, childhood test scores, parental education, household assets, and age are measured in 2006, when respondents were around age 12; gender and urban residence are measured in 2002. The urban coefficient is blank for Pakistan because all respondents are rural and for Ethiopia and India because urban residence is collinear with the sampling-cluster fixed effects. The p-value is from a joint test that all predictors included in the corresponding multiple regression equal zero. All specifications include district fixed effects for Pakistan and sampling-cluster fixed effects for the Young Lives countries. Standard errors are clustered at the village level in Pakistan and the community level in the Young Lives countries. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Correlation with Years of Schooling

	LEAPS (24.5y)	Young Lives (28.5y)		
	(1)	(2)	(3)	(4)
	Pakistan	Ethiopia	India	Peru
Panel A: Full Sample				
Child IRT Score	1.543*** (0.126)	1.375*** (0.195)	1.155*** (0.120)	1.122*** (0.139)
Sample Mean	9.402	10.602	12.245	12.959
Adjusted R ²	0.383	0.294	0.310	0.286
Observations	955	610	801	486
Clusters	108	24	99	59
Panel B: Females Only				
Child IRT Score	1.659*** (0.206)	1.213*** (0.318)	1.141*** (0.183)	1.207*** (0.209)
Sample Mean	9.806	10.793	11.713	12.879
Adjusted R ²	0.410	0.312	0.296	0.343
Observations	428	290	407	231
Clusters	106	24	97	52
Panel C: Males Only				
Child IRT Score	1.453*** (0.158)	1.503*** (0.230)	1.090*** (0.145)	1.011*** (0.234)
Sample Mean	9.074	10.428	12.794	13.031
Adjusted R ²	0.357	0.287	0.309	0.212
Observations	527	320	394	255
Clusters	105	22	94	48

Notes: This table reports the association between childhood test scores and years of schooling by the latest survey round. Column (1) reports the estimate for Pakistan, while Columns (2)–(4) report estimates for the Young Lives older cohort in Ethiopia, India, and Peru, respectively. In LEAPS, childhood test scores are the average of mathematics, Urdu, and English IRT scores from tests administered between 2003 and 2006, and the child-weighted mean testing age is 11.2 years. In Young Lives, childhood test scores are mathematics IRT scores from 2006, when respondents were around age 12. LEAPS scores are standardized using the LEAPS analytic-sample distribution, while Young Lives scores are standardized using the Indian older-cohort analytic-sample distribution. Because LEAPS and Young Lives do not share linking items, test-score levels are not directly comparable between the two studies. All specifications control for baseline household wealth, measured using a within-country standardized asset index; parental education, measured as indicators for whether each parent has any formal education; age fixed effects; and district fixed effects for LEAPS or sampling-cluster fixed effects for Young Lives. Full-sample specifications additionally include gender fixed effects. Standard errors are clustered at the village level for LEAPS and at the community level for Young Lives. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Correlation with College or Vocational Education Attendance

	LEAPS (24.5y)	Young Lives (28.5y)		
	(1)	(2)	(3)	(4)
	Pakistan	Ethiopia	India	Peru
Panel A: Full Sample				
Child IRT Score	0.093*** (0.012)	0.105*** (0.025)	0.135*** (0.017)	0.128*** (0.020)
Sample Mean	0.161	0.425	0.559	0.679
Adjusted R ²	0.189	0.212	0.245	0.173
Observations	955	610	801	486
Clusters	108	24	99	59
Panel B: Females Only				
Child IRT Score	0.095*** (0.022)	0.102** (0.043)	0.113*** (0.027)	0.124*** (0.028)
Sample Mean	0.206	0.448	0.467	0.662
Adjusted R ²	0.216	0.274	0.191	0.193
Observations	428	290	407	231
Clusters	106	24	97	52
Panel C: Males Only				
Child IRT Score	0.089*** (0.014)	0.121*** (0.023)	0.150*** (0.023)	0.119*** (0.038)
Sample Mean	0.125	0.403	0.655	0.694
Adjusted R ²	0.186	0.171	0.258	0.142
Observations	527	320	394	255
Clusters	105	22	94	48

Notes: This table reports the association between childhood test scores and whether the respondent has ever attended college/vocational education by the latest survey round. Column (1) reports the estimate for Pakistan, while Columns (2)–(4) report estimates for the Young Lives older cohort in Ethiopia, India, and Peru, respectively. In LEAPS, childhood test scores are the average of mathematics, Urdu, and English IRT scores from tests administered between 2003 and 2006, and the child-weighted mean testing age is 11.2 years. In Young Lives, childhood test scores are mathematics IRT scores from 2006, when respondents were around age 12. LEAPS scores are standardized using the LEAPS analytic-sample distribution, while Young Lives scores are standardized using the Indian older-cohort analytic-sample distribution. Because LEAPS and Young Lives do not share linking items, test-score levels are not directly comparable between the two studies. All specifications control for baseline household wealth, measured using a within-country standardized asset index; parental education, measured as indicators for whether each parent has any formal education; age fixed effects; and district fixed effects for LEAPS or sampling-cluster fixed effects for Young Lives. Full-sample specifications additionally include gender fixed effects. Standard errors are clustered at the village level for LEAPS and at the community level for Young Lives. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Correlation with High School Completion

	LEAPS (24.5y)	Young Lives (28.5y)		
	(1)	(2)	(3)	(4)
	Pakistan	Ethiopia	India	Peru
Panel A: Full Sample				
Child IRT Score	0.134*** (0.013)	0.097*** (0.019)	0.140*** (0.017)	0.076*** (0.022)
Sample Mean	0.254	0.228	0.610	0.870
Adjusted R ²	0.256	0.118	0.195	0.153
Observations	955	610	801	486
Clusters	108	24	99	59
Panel B: Females Only				
Child IRT Score	0.147*** (0.024)	0.098** (0.037)	0.141*** (0.026)	0.086** (0.037)
Sample Mean	0.315	0.217	0.568	0.853
Adjusted R ²	0.295	0.214	0.222	0.142
Observations	428	290	407	231
Clusters	106	24	97	52
Panel C: Males Only				
Child IRT Score	0.122*** (0.015)	0.087*** (0.029)	0.134*** (0.023)	0.069** (0.028)
Sample Mean	0.205	0.237	0.655	0.886
Adjusted R ²	0.208	0.083	0.165	0.116
Observations	527	320	394	255
Clusters	105	22	94	48

Notes: This table reports the association between childhood test scores and high school completion by the latest survey round. Column (1) reports the estimate for Pakistan, while Columns (2)–(4) report estimates for the Young Lives older cohort in Ethiopia, India, and Peru, respectively. In LEAPS, childhood test scores are the average of mathematics, Urdu, and English IRT scores from tests administered between 2003 and 2006, and the child-weighted mean testing age is 11.2 years. In Young Lives, childhood test scores are mathematics IRT scores from 2006, when respondents were around age 12. LEAPS scores are standardized using the LEAPS analytic-sample distribution, while Young Lives scores are standardized using the Indian older-cohort analytic-sample distribution. Because LEAPS and Young Lives do not share linking items, test-score levels are not directly comparable between the two studies. All specifications control for baseline household wealth, measured using a within-country standardized asset index; parental education, measured as indicators for whether each parent has any formal education; age fixed effects; and district fixed effects for LEAPS or sampling-cluster fixed effects for Young Lives. Full-sample specifications additionally include gender fixed effects. Standard errors are clustered at the village level for LEAPS and at the community level for Young Lives. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Predictive Decomposition Components (Men Only)

	LEAPS (24.5y)	Young Lives (28.5y)		
	(1)	(2)	(3)	(4)
	Pakistan	Ethiopia	India	Peru
Panel A: No SES controls				
Eq. (2): Total gradient, $\hat{\pi}$	0.121	0.019	0.220	0.216
Eq. (3): Adult score-earnings, $\hat{\beta}$	0.130	0.031	0.249	0.097
Eq. (4): Schooling-earnings, $\hat{\theta}$	0.137	0.096	0.220	0.219
Eq. (5): Child score-adult score, $\hat{\gamma}$	0.560	0.376	0.493	0.348
Eq. (6): Child score-schooling, $\hat{\delta}$	0.464	0.391	0.439	0.224
Eq. (7): Corr(adult score, schooling), $\hat{\rho}$	0.717	0.582	0.490	0.407
Eq. (8): Predicted component, $\hat{\pi}^{\text{expl}}$	0.079	0.032	0.148	0.051
Eq. (9): Residual component, $\hat{\pi}^{\text{res}}$	0.042	-0.013	0.071	0.164
Observations	335	198	323	224
Panel B: With SES controls				
Eq. (2): Total gradient, $\hat{\pi}$	0.098	0.009	0.134	0.203
Eq. (3): Adult score-earnings, $\hat{\beta}$	0.085	0.029	0.147	0.072
Eq. (4): Schooling-earnings, $\hat{\theta}$	0.079	0.094	0.121	0.190
Eq. (5): Child score-adult score, $\hat{\gamma}$	0.546	0.376	0.414	0.324
Eq. (6): Child score-schooling, $\hat{\delta}$	0.447	0.384	0.357	0.185
Eq. (7): Corr(adult score, schooling), $\hat{\rho}$	0.682	0.582	0.391	0.343
Eq. (8): Predicted component, $\hat{\pi}^{\text{expl}}$	0.049	0.030	0.076	0.037
Eq. (9): Residual component, $\hat{\pi}^{\text{res}}$	0.049	-0.021	0.059	0.166
Observations	335	198	323	224

Notes: This table reports the component estimates underlying the predictive decomposition in Table 2. Columns (1)–(4) report estimates for Pakistan and the Young Lives older cohort in Ethiopia, India, and Peru, respectively. The sample is restricted to men with positive earnings and non-missing log earnings, childhood test scores, adult test scores, years of schooling, baseline SES controls, age, and district or sampling-cluster identifiers. Earnings are converted to 2023 PPP dollars before being logged. Within each country and panel, log earnings, childhood test scores, adult test scores, and years of schooling are residualized and standardized before the decomposition. Panel A includes age fixed effects and district fixed effects for LEAPS or sampling-cluster fixed effects for Young Lives. Panel B additionally controls for baseline household wealth, measured using a within-country standardized asset index, and parental education, measured as indicators for whether each parent has any formal education. Eq. (2), $\hat{\pi}$, is the total child-score earnings gradient. Eq. (3), $\hat{\beta}$, is the adult-score earnings gradient. Eq. (4), $\hat{\theta}$, is the schooling earnings gradient. Eq. (5), $\hat{\gamma}$, is the child-score adult-score gradient. Eq. (6), $\hat{\delta}$, is the child-score schooling gradient. Eq. (7), $\hat{\rho}$, is the correlation between adult test scores and schooling. Eq. (8), $\hat{\pi}^{\text{expl}}$, is the predicted component implied by the adult-test-score and schooling pathways, allowing adult test scores and schooling to be correlated. Eq. (9), $\hat{\pi}^{\text{res}}$, is the residual component, equal to Eq. (2) minus Eq. (8). Because all variables entering the decomposition are standardized within country and panel, the reported gradients and components are in standard-deviation units.

Appendix B: Data Sources and Measurement

B.1 LEAPS

B.1.1 Study Design and Sample

The Pakistan data come from Learning and Educational Achievement in Pakistan Schools (LEAPS), a longitudinal study of education markets in 112 villages across three districts of Punjab province—Attock, Faisalabad, and Rahim Yar Khan—that had at least one private school in 2000 (Andrabi et al., 2007). Because of this restriction, the sampled villages were wealthier than the average village but representative of the types of villages in which approximately 60 percent of the population lived (Andrabi et al., 2022). Within sampled villages, the LEAPS team surveyed all primary schools in the village or within a feasible walking distance in 2003. They also independently sampled 1,807 households (~ 16 per village) from the village household listing that year. The household sample was stratified by children’s grade-3 enrollment status: in each village, 75% of target households were sampled from households with at least one child enrolled in grade three, while the rest were sampled from households with an age-eligible child not enrolled in grade three.

As shown in Appendix Figure B1, the team re-visited the 112 villages annually between 2004 and 2006, and again in 2011, surveying all schools in each village and following the original panel of 1,807 households in each round. The team then completed a fifth follow-up round in 2018/19. Because many of the original households had split by this time, the project tracked individuals, rather than households, and attempted to re-contact all 5,865 children aged 5–15 in 2003, regardless of their location. The tracking rate was 84.5%, implying an annual attrition rate slightly above 1% (Danon et al., 2024).

In this paper, we focus on individuals in the original household sample who were tested at least once between 2003 and 2006, have non-missing baseline controls, and were re-observed in 2018/19. Appendix Table B1 shows that these restrictions reduce the sample from 1,269 children with a childhood test score to 1,119 with non-missing baseline controls, and, finally, to an analytic sample of 955 individuals observed in 2018/19.

B.1.2 Childhood Test Scores and Baseline Controls

Between 2003 and 2011, the LEAPS team independently administered low-stakes tests in English, Urdu, and Mathematics to several cohorts of children. Our measure of foundational learning uses the school-based assessments administered between 2003 and 2006. In 2003, the LEAPS team tested a first cohort of grade-3 children. These children were re-tested in school in subsequent rounds (2004, 2005, and 2006), regardless of their grade. In 2005, they also tested a new cohort of children in grade 3 and then tested them again in 2006. Finally, in 2011, the team tested a

new cohort of grade 4 children in school. They also expanded the surveys to include home-based tests for members of the two earlier cohorts and one randomly selected sibling between ages 5 and 15. We do not use the 2011 assessments to construct our measure of foundational learning. [Andrabi et al. \(2002\)](#) provide a detailed discussion of the assessments and their validity.

Test scores were constructed using Item Response Theory (IRT). Using repeated anchor items, we estimate subject-specific IRT models and recover expected a posteriori (EAP) ability estimates on vertically linked scales that are comparable across the 2003–2011 rounds. Across the pooled test-taking population, each subject-specific EAP score has a mean of ~ 0 and a standard deviation of ~ 1 . To construct our measure of foundational learning, we first average the available subject-specific scores within each child-round. We then average these round-specific scores across all rounds in which the child was tested. Finally, we standardize the resulting measure to have a mean of zero and a standard deviation of one in the analytic sample.

In our analytic sample of 955 children, 51% were tested in all four rounds, 19.4% in three rounds, 23.6% in two rounds, and 6.1% in one round.¹⁶ Most children were ages 7–12 in 2003, with a median age of 10. The child-weighted mean testing age was 11.2 years, so the measure captures learning near the end of primary school.

Finally, we use three baseline socioeconomic controls: mother’s education, father’s education, and household wealth. The first two indicate whether the child’s mother and father have any formal education. Because parental education was reported in multiple rounds, we construct these indicators using a majority rule across the 2003–2006 rounds. If parental education is not observed in any of these rounds, we use the reported levels from the 2011 survey. We measure household wealth using an asset index, constructed as the first principal component of 25 asset-ownership indicators measured in 2003. We standardize the index with respect to the unique households represented in our analytic sample.

B.1.3 Long-Term Follow-Up and Outcomes

We measure long-term outcomes in 2018/19, approximately 15 years after the baseline round (2003). Of the 1,119 individuals with baseline test scores and non-missing SES controls, we re-observed 955 (85.3%) in 2018/19 (Appendix Table B1). Appendix Table A1 shows neither childhood test scores nor baseline SES or demographic characteristics predict observation at endline, either individually or jointly (p -value = 0.894). Among the 955 observed, 842 were interviewed in person (88.2%) and 16 by phone. For the remaining 97, information was provided by another household member. Information on

¹⁶The distribution is similar for the full sample of 1,269 test-takers: 50.3% in all four rounds, 19.5% in three rounds, 23.1% in two rounds, and 7.1% in one round.

educational attainment and labor-market outcomes was collected in all three survey modes, while adult skills assessments were only administered to those interviewed in person.

We construct three sets of long-term outcomes: adult skills, educational attainment, and labor market outcomes.

Adult Skills: The LEAPS team administered cognitive assessments in English, Urdu, and Mathematics to in-person respondents, who were 24.5 years old, on average. Although these tests retained common items with tests from 2003–2011, we do not impose inter-year linkage between childhood and adult test scores. Successful intertemporal linkage requires stability in item response functions, conditional on latent ability. By 2018/19, many children had been out of school for several years. This pattern likely reinforced frequently used practical skills while depreciating less-used academic skills, which potentially induces differential item functioning. We therefore estimate subject-specific IRT models using the 2018/19 test-taking population and recover EAP ability estimates on adult-specific scales.¹⁷ As with childhood scores, we average subject-specific scores for each respondent to construct a single measure of adult skills and standardize this measure among test-takers in our analytic sample (840 individuals or 88% of the sample).

Educational Attainment: We construct four measures of educational attainment from the 2018/19 survey. Years of schooling is based on the respondent’s current grade/program if enrolled and highest completed grade otherwise. We convert vocational, technical, and university programs into equivalent years of education based on the program and the respondent’s enrollment status. Current enrollment indicates whether the respondent was enrolled in an educational program (usually tertiary programs given their ages) at the time of the survey. High school completion indicates whether the individual has completed upper-secondary school (grade 12). College or vocational attendance indicates whether the respondent was currently enrolled in or had previously completed a vocational, technical, polytechnic, or university program.

Labor Market: We construct two labor market outcomes: work status and monthly earnings. For Pakistan, work status indicates whether the respondent was engaged in any paid activity in the past month. We construct our earnings measure from the respondent’s primary and secondary activities, which include daily wage work, salaried employment, self-employment, family businesses, and agriculture.

For those engaged in daily wage work, self-employed work, or family business, we average their monthly earnings over the past two months. For salaried workers, we aggregate their monthly earnings across jobs. We construct agricultural earnings in two ways. For in-person

¹⁷The 2018/19 test-taking population included 4,398 adults, 88.7% of the 4,956 individuals observed in the round.

respondents who list agriculture as their primary or secondary activity, we compute monthly net agricultural earnings as the difference between annual crop and livestock revenues and the full set of reported labor and input costs, divided by 12. When these computed agricultural earnings are unavailable, we use the respondent’s estimated agricultural earnings. We also use estimated agricultural earnings for respondents who do not list agriculture as their primary or secondary activity but report engaging in agricultural work with the intention of selling their produce for more than four weeks in the past year. For women interviewed in person, we additionally capture informal earnings from tutoring or teaching, sewing, and other paid activities. We then calculate total monthly earnings by aggregating earnings from each source.

We convert earnings to 2023 PPP US dollars to facilitate cross-country comparisons. In the regression analyses presented in Section 3, we use log monthly earnings and restrict the sample to respondents with positive earnings. For the development accounting exercise (Section 4), which uses earnings in levels, we winsorize earnings among workers at the 1st and 99th percentiles within country-cohort-gender cells.

B.2 Young Lives

B.2.1 Study Design and Sample

The Young Lives study is a longitudinal study that has tracked two cohorts of children, an “older cohort” born in 1994/95 and a “younger cohort” born in 2001/02, in Ethiopia, India (Andhra Pradesh and Telangana), and Peru from 2002 to 2023/24.¹⁸ In each country, 20 sentinel sites (clusters) were selected (woredas in Ethiopia, mandals in India, and districts in Peru), and approximately 100 and 50 participants were randomly sampled in each cluster for the younger and older cohorts. These samples are not statistically representative of the countries but, in comparison with Demographic and Health Surveys, show substantial variation in living standards (Outes-Leon and Sanchez, 2008; Kumra, 2008; Escobal and Flores, 2008).

As shown in Appendix Figure B2, Young Lives revisited both cohorts six times after the 2002 baseline, completing the most recent round in 2023/24. Members of the older and younger cohorts were, on average, 28.5 and 21.5 years old, respectively, during the latest round. Because many members of the younger cohort remained enrolled in education, their transitions from school to work were not yet complete. We therefore focus our analysis on the older cohort.

As with Pakistan, we restrict the older-cohort sample to respondents who have childhood test scores and non-missing baseline SES controls and who were re-observed in 2023/24. Appendix Table B1 shows that these restrictions reduce the sample from 2,542 children with a childhood test score to 2,428 with non-missing baseline controls, and, finally, to an analytic sample of 1,897 individuals observed in 2023/24: 610 respondents in Ethiopia, 801 in India, and 486 in Peru.

B.2.2 Childhood Test Scores and Baseline Controls

Across survey rounds, the Young Lives team consistently administered independent, low-stakes Mathematics and Language tests to both cohorts. For cross-country comparability, we focus on Mathematics scores in this paper. We observe test scores for the older cohort between Rounds 2 and 4 (2006–2013) and for the younger cohort between Rounds 3 and 5. Our measure of foundational learning uses the assessments administered to the older cohort in 2006, when respondents were 11.7 years old, on average.

As in Pakistan, test scores were constructed using the IRT methodology. We use the scores publicly released by the Young Lives team (Espinoza Revollo and Scott, 2022). Using repeated anchor items, the Young Lives team jointly estimated a two-parameter IRT model across countries, cohorts, and rounds and recovered EAP ability estimates on a linked scale. The released scores were normalized to have a mean of 500 and a standard deviation of 100

¹⁸The Young Lives study also followed these two cohorts in Vietnam. However, the survey was not fielded in Vietnam in 2023/24 due to a change in governmental procedures for the international transfer of personal data. Since we focus on outcomes measured in the latest round, we exclude Vietnam from our analysis.

among round two test-takers. For our analysis, we standardize scores using the Indian older-cohort analytic sample distribution and apply the same transformation to Ethiopia and Peru. After this standardization, mean childhood test scores are 0 in India, -0.64 in Ethiopia, and -0.03 in Peru, expressed in standard deviations of the Indian older-cohort analytic-sample distribution.¹⁹ This common transformation preserves the cross-country differences in score levels used in the development accounting exercise in Section 4.

Finally, as in Pakistan, we use three baseline socioeconomic controls: mother’s education, father’s education, and household wealth, all measured in 2006. The first two indicate whether the respondent’s mother and father have any formal education. We measure household wealth using a country-specific asset index constructed as the first principal component of 21 indicators of household durables and housing characteristics in Ethiopia and India and 31 indicators in Peru. We construct the index separately for each country and standardize it using the corresponding country’s analytic-sample distribution.

B.2.3 Long-Term Follow-Up and Outcomes

We measure educational attainment and labor market outcomes in 2023/24, approximately 17 years after the 2006 baseline. Adult skills are measured in 2013, the last year for which we have test scores for the older cohort. Of the 2,428 respondents with childhood test scores and non-missing baseline SES controls, 1,897 (78.1%) were re-observed in 2023/24 (Appendix Table B1). Re-observation rates were 71.8% in Ethiopia, 85.7% in India, and 75.6% in Peru. Appendix Table A1 shows that childhood test scores do not predict observation at endline in any country and that baseline characteristics do not jointly predict observation in any of the three countries (p -values = 0.143, 0.670, and 0.287, respectively). Interviews were conducted in person, except for 58 respondents in Ethiopia who completed a reduced phone survey because conflict prevented in-person fieldwork in two sites in the Amhara region.

We construct three sets of long-term outcomes: adult skills, educational attainment, and labor market outcomes.

Adult Skills: The Young Lives team administered an independent, low-stakes mathematics assessment to the older cohort in 2013, when respondents were 18.6 years old, on average. We use the publicly released IRT scores described above, which place the 2006 and 2013 assessments on a linked scale. For our analysis, we standardize the adult scores using the Indian older-cohort analytic sample distribution and apply the same transformation to Ethiopia and Peru. We observe adult test scores for 1,850 of the 1,897 respondents in our analytic sample (97.5%).

Educational Attainment: We construct four measures of educational attainment. For years of schooling, high school completion, and college or vocational attendance, we combine education

¹⁹The corresponding standard deviations are 1.00 in India, 0.96 in Ethiopia, and 0.91 in Peru.

histories collected in 2016 and 2023/24, using earlier rounds (2009 and 2012) to fill missing 2016 values for a subset of respondents, and retain the highest reported attainment. Years of schooling is based on the highest grade completed in formal education, or the current grade for respondents who remain enrolled, with technical, vocational, and university programs converted into equivalent years using country-specific program structures. Current enrollment indicates whether the respondent was enrolled in formal education in 2023/24. High school completion indicates whether the respondent completed upper-secondary school (grade 12 in Ethiopia and India and grade 11 in Peru). College or vocational attendance indicates whether the respondent ever attended a technical, vocational, or university program.

Labor Market: We construct two labor-market outcomes: work status and monthly earnings. Work status indicates whether the respondent worked for at least one hour in the past week on a household farm, for someone outside the household, or in a self- or family-owned enterprise, regardless of whether they were paid. This measure excludes studying, housework, and unpaid care for household members.

For respondents who worked during the previous week, the survey identified their main activity as the paid or unpaid activity on which they spent the most time. Respondents who were temporarily absent from a current job reported information about that job. Respondents separately reported net earnings from their main activity in cash and in kind. Net earnings include wages or salaries, tips, gratuities, bonuses, and the value of in-kind payments after deducting taxes and other work-related expenses; respondents working in their own businesses were instructed to deduct production costs. Extra employment benefits are excluded. Unlike in Pakistan, the earnings measure does not incorporate earnings from secondary activities.

We convert reported earnings into monthly amounts based on the period covered by the payment. We multiply hourly, daily, weekly, and piece-rate earnings by the respondent's reported hours per day, days per week, weeks per month, or pieces produced per day, as applicable. We retain monthly earnings as reported, divide annual earnings by the number of months worked during the year, and multiply fortnightly earnings by two. We then sum cash and in-kind earnings. We assign zero earnings to respondents who worked but reported receiving no payment and to non-workers whose earnings were otherwise missing.²⁰

We convert earnings to 2023 PPP US dollars to facilitate cross-country comparisons.²¹ In the regression analyses presented in Section 3, we use log monthly earnings and restrict the sample to respondents with positive earnings. For the development accounting

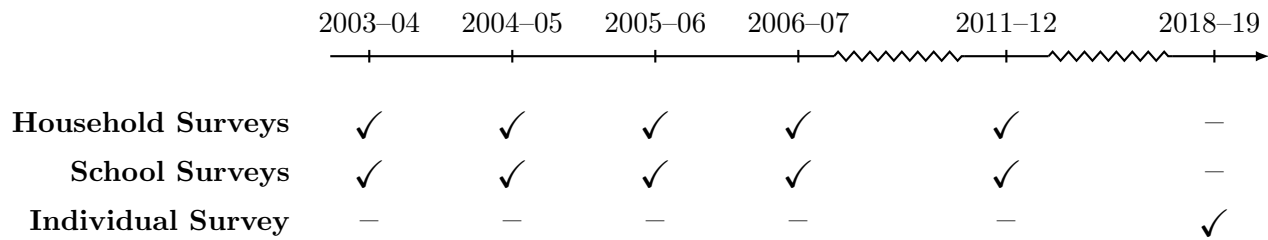
²⁰Thirty respondents—11 in Ethiopia, 11 in India, and eight in Peru—reported having a current job despite not working during the previous week. We code them as not working under our seven-day work-status definition but retain the positive earnings reported for their current job.

²¹For respondents interviewed in Ethiopia in 2024, we first deflate earnings to 2023 local-currency values using the Ethiopian consumer price index and then apply the 2023 PPP conversion factor.

exercise in Section 4, which uses earnings in levels, we winsorize earnings among workers at the 1st and 99th percentiles within country-cohort-gender cells.

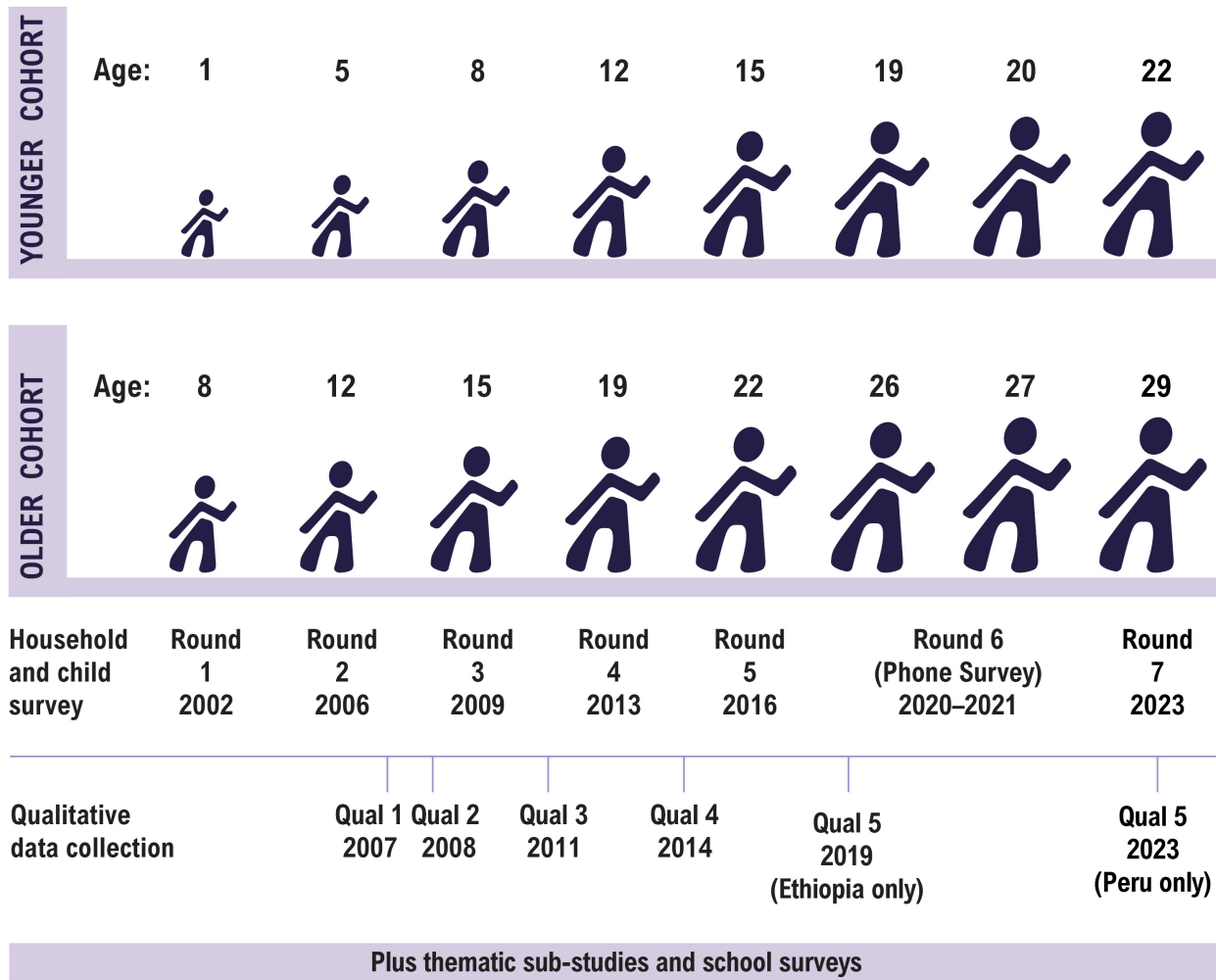
B.3 Figures

Figure B1: Timeline of LEAPS Data Collection



Notes: Check marks indicate the rounds in which each type of survey was administered; dashes indicate that the survey was not administered. In the 2003-04 through 2006-07 rounds and again in 2011-12, LEAPS surveyed schools and followed the original panel of 1,807 households in the 112 study villages. In 2018-19, LEAPS shifted to an individual survey that attempted to track all children who were ages 5-15 in the 2003 household roster, regardless of their location.

Figure B2: Timeline of Young Lives Data Collection



Notes: Ages shown are approximate. Our analysis uses the older cohort, measuring childhood test scores in Round 2 (2006), adult skills in Round 4 (2013), and educational attainment and labor-market outcomes in Round 7 (2023/24). The qualitative data collections, thematic sub-studies, and school surveys shown in the figure are not used in this paper.

B.4 Tables

Table B1: Construction of Analytic Samples

	Has Childhood Test Scores	+ Non-Missing SES Controls	+ Observed at Endline	Endline Re-observation Rate (%)
Pakistan	1,269	1,119	955	85.3
Ethiopia	919	850	610	71.8
India	943	935	801	85.7
Peru	680	643	486	75.6

Notes: The first three columns report cumulative sample restrictions. Childhood test scores refer to respondents with at least one assessment between 2003 and 2006 in Pakistan and those with a mathematics assessment in 2006 in Ethiopia, India, and Peru. SES controls comprise mother’s education, father’s education, and baseline household wealth. Endline refers to the 2018/19 survey in Pakistan and the 2023/24 survey in Ethiopia, India, and Peru. The endline re-observation rate is calculated among respondents with childhood test scores and non-missing SES controls.