CENTER FOR GLOBAL DEVELOPMENT

The Long-Run Decline of Education Quality in the Developing World

Alexis Le Nestour, Laura Moscoviz, Justin Sandefur

Abstract

We use comparable, survey-based literacy tests for repeated cross-sections of men and women born between 1950 and 2000 to study education outcomes across cohorts in 87 countries. We find that education quality, defined as literacy conditional on completing five years of schooling, stagnated across the developing world over half a century, including absolute declines in both South Asia and sub-Saharan Africa. Shifts in student composition clearly explain part of the downward trend we observe, but the decline pre-dates the abolition of school fees in most countries, and anthropometric data suggest students in later decades were healthier and wealthier than those in earlier cohorts. Globally, increases in schooling outpaced the decline in education quality, leading to a large increase in unconditional literacy.

KEYWORDS
literacy, education quality, access to education

JEL CODES
I25, N37, O15

This paper was originally published in February 2022. It was updated in September 2023. The original version can be viewed here.

WORKING PAPER 608 · SEPTEMBER 2023

The Long-Run Decline of Education Quality in the Developing World

Alexis Le Nestour
UNICEF Office of Research (Innocenti)

Laura Moscoviz
Center for Global Development

Justin Sandefur
Center for Global Development

Replication materials and country-by-country results are available at the hyperlinks. This project benefited from discussions with Girin Beeharry, Asyia Kazmi, and Lant Pritchett, and comments from Maryam Akmal, Lee Crawfurd, Jean Dreze, David Evans, Eric Hanushek, Susannah Hares, Kim Lehrer, Isaac Mbiti, Jack Rossiter, Arvind Subramanian, Francis Teal, Andrew Zeitlin, and seminar participants at RISE and CSAE Oxford. Thi Le provided excellent research assistance. This work was supported by a grant from the Bill & Melinda Gates Foundation. The views expressed here should not be attributed to the Center for Global Development or its funders. All remaining errors are our own.

Alexis Le Nestour, Laura Moscoviz, and Justin Sandefur. 2023. "The Long-Run Decline of Education Quality in the Developing World." CGD Working Paper 608. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/long-run-decline-education-quality-developing-world

CENTER FOR GLOBAL DEVELOPMENT
2055 L Street, NW Fifth Floor
Washington, DC 20036
202.416.4000

1 Abbey Gardens
Great College Street
London
SW1P 3SE

www.cgdev.org

1 Introduction

International organizations like the World Bank and UNESCO have declared a "learning crisis" in the developing world, with many school systems failing to reliably produce even basic literacy and numeracy skills [World Bank, 2018, UNESCO, 2013]. Pupils in India, for instance, score on average at about the 5th percentile for pupils at a similar grade in advanced economies on international learning assessments, while nearly 80 percent of South African children cannot read for meaning in any language by fourth grade [Das and Zajonc, 2010, Howie et al., 2017]. The economic consequences of improving education quality are poorly understood, but potentially large. Recent work in the macroeconomics of development has suggested that education quality may outperform quantity measures in explaining long-run growth in econometric tests, and accounting for quality differentials roughly doubles the explanatory power of education in a development accounting framework [Hanushek and Kimko, 2000, Hanushek and Wossmann, 2006, Schoellman, 2012].[1]

The idea of a learning "crisis" implies something new. But we lack reliable, long-term measures of education quality over time, particularly for the developing world, to evaluate competing explanations for this crisis.[2] For instance, one prominent narrative is that the expansion of mass schooling, e.g., after the abolition of user fees in many African countries in the 1990s and 2000s, led to a quality decline [Ruto and Mugo, 2005, Pritchett, 2013, Taylor and Spaull, 2013, Atuhurra, 2016].[3] Another complementary narrative is that countries at the top of the league table in international learning assessments (e.g., Finland or Vietnam) hold reform lessons for their neighbors [Sahlberg, 2007, Dang and Glewwe, 2018]. This argument often attributes current success to current educational policies, and elides the question of timing. Is, say, Vietnam's success a feature of reforms to its education system during its recent liberalization period, or a longer-standing national characteristic? Are the

[1] This view is not universal: Caselli [2005] questions the explanatory power of test scores in development accounting, based on the weak observed Mincerian returns to test scores for individual incomes in U.S. micro data. Kraay [2018] suggests a different approach, eschewing the Mincerian literature, and attributing a bigger role to education quality.

[2] While a number of developing countries have participated in multiple rounds of international or regional learning assessments, the time span covered is generally limited. For instance, 14 African countries participated in two rounds of the SACMEQ study from 2000 to 2007 while the PASEC, which covers 18 African countries, started in 1995 but has no strict comparability across countries before 2014 and no comparable time series. 19 Latin American countries participated in the Laboratorio SERCE and TERCE tests from 1997 to 2019, and 41 developing countries participated in multiple rounds of TIMSS, PIRLS, or PISA, generally spanning just a single decade or less.

[3] Though see also Blimpo et al. [2019] for a contradictory finding from The Gambia.

lagging learning outcomes in India, or the truly abysmal outcomes we document in Nigeria, the result of a contemporary crisis or do they have deeper historical roots?

To answer these questions, we construct new, comparable time series of education quality for 87 developing countries, stretching from cohorts born in the 1950s up to the late 1990s. We define education quality as the literacy rate for an adult with a given level of schooling, typically five years of primary school. This is roughly analogous to major international learning assessments such as TIMSS, PIRLS, etc., which measure education quality with standardized tests in a specified grade. To estimate literacy across age cohorts, we draw on repeated cross-sections from the Demographic and Health Surveys (DHS) run by USAID and Multiple Indicator Cluster Surveys (MICS) run by UNICEF, which provide direct, rather than self-reported, tests of literacy for adult women and, in a subset of survey rounds, men, aged 15 to 49, mostly measured between the early 2000s and late 2010s. We also disaggregate results for some larger countries, i.e., India, Nigeria, and Indonesia, to look for patterns of divergence or convergence among sub-national units.

The core analytical challenge throughout the paper is to identify cohort effects in repeated cross-sectional data. As is well known, age, period, and cohort effects cannot be separately identified in a simple linear model without invoking auxiliary assumptions about the underlying data generating process [Fosse and Winship, 2019]. To achieve identification, we follow Deaton and Paxson [1994] and impose the restriction that period effects sum to zero and are orthogonal to a trend. This method has been widely used in the economics literature dealing with age, period and cohort models (e.g., by Lagakos et al. [2018] to estimate experience-wage profiles). In our setting, the substantive content of the Deaton-Paxson normalization is that we allow for literacy to accumulate or decay over an individual's life cycle, and to vary from one cohort to another, but rule out sustained periods where the adult population acquires or loses literacy en masse. This assumption is corroborated by the fact individuals with zero schooling also show nearly zero literacy in our data with no upward trend over time. In short, we interpret systematic differences between surveys that cannot be accounted for by age or cohort effects as survey effects, i.e., artifacts of data collection rather than facts about the world.

Results from our age-period-cohort model of literacy rates reveal a picture of overall stagnation in education quality in the developing world, and fairly stable gaps in quality across countries. We find virtually no case worldwide of dramatic improvements in education quality over a fifty-year time horizon, and for most of the 87 countries in our sample we see flat lines, with little movement up or down. This flat trend holds over multiple decades for some notable high-performers like Burundi and Vietnam, and low-performers like Nigeria. However, we also find evidence of significant secular decline in education quality for a small number of large developing countries, notably India and Ethiopia. The Indian data are somewhat inconsistent across rounds, however, which we discuss below and in a separate appendix.

Changing patterns of selection into school are an obvious potential explanation for declines in measured quality, which we explore in Section 6. This is particularly salient given the dramatic increase in average years of schooling observed in most countries in our sample over this time frame, and the possibility that marginal students possessed lower levels of non-school inputs in the human capital production function (e.g., worse nutrition). Indeed, we find that on average the selection of pupils into schooling on the basis of height, a characteristic largely determined prior to schooling decisions, weakens over time. Furthermore, in countries and time periods where enrollment expanded more rapidly, education quality declined more quickly, consistent with (though clearly not definitive proof of) a selection effect. Controlling for pupil height or local enrollment trends dampens the measured downward trend in education quality in our overall sample, but leaves a significant negative trend for both men and women. Furthermore, average adult height among individuals who went to school is increasing over this period, implying that entering pupils may have been better prepared in absolute terms even if they comprised a less select group in relative terms. We also find, on average, no significant trend break in observed education quality when countries abolished user fees for primary schooling, a policy change which putatively expanded access to schooling for marginal pupils.

Our results add a new dimension to the large literature measuring human capital around the world (i.e., the evolution of quality over time), with relevance for a variety of empirical applications. The most widely used human capital measure in the economics, based on Barro and Lee [1993] and subsequent updates, focused initially on quantity rather than quality. More recent work has incorporated richer measures of education quality, but this has come at the cost of either limited coverage for the developing world, or short time series [Yue Yin, 2018, Sulis et al., 2020, Gustafsson, 2016, Van Damme and Bellens, 2017, Kraay, 2018, Patel and Sandefur, 2020]. [4]By combining a quality dimension with long time series and broad

[4]The World Bank's Harmonized Learning Outcomes, based on early work by Altinok and Murselli [2007] and extended and refined in Altinok et al. [2018], is an important instance of this approach, which covers a large number of developing countries and reports figures from 1965 onward. Note, however, that the actual

developing country coverage, the results here would potentially enable future work to revisit earlier cross-country findings on the determinants of test-score performance, and conversely, the role of education quality in promoting economic development [Hanushek and Wossmann, 2006, Wossmann and Hanushek, 2007]. As an illustration, we find a strong correlation of quality with subsequent economic growth in our developing country sample, though this finding is not robust to controlling for unobserved country effects.

The following section describes the coverage of our survey data sources in more detail and in Section 3 we describe and present some basic validation tests of our measure of education quality. We then turn to our core empirical strategy to overcome the challenge of disentangling age, period, and cohort effects in repeated cross-sectional data on literacy in Section 4. Having found mostly downward trends in observed quality, Section 6 then examines whether changing patterns of selection into schooling drive these trends. Finally, Section 5 presents various stylized facts based on the trends in education quality that we calculate, and section 7 concludes.

2 Data

Our dataset consists in the universe of Demographic and Health (DHS) and Multiple Indicator Cluster surveys (MICS) as of January 2021 with information on literacy and schooling. DHS and MICS collect nationally representative information on literacy and schooling of women, and in some cases men, aged 15 to 49.[5] Since 2000, reading abilities are tested with a short literacy test consisting on reading four simple sentences out loud. Interviewers then score individuals on a three point scale: "cannot read at all", "able to read only parts of the sentence" and "able to read whole sentence". We use a 0-1 scale for literacy with individuals "able to read only parts of the sentence" scoring 0.5 point. Prior to 2000, literacy was self-reported by individuals on a similar three point scale and surveys with self-reported data on literacy are excluded to keep a consistent measure of literacy over time.

testing data before 1990 covers only 23, mostly advanced economies, increasingly gradually thereafter. Thus in more recent applications the World Bank has restricted focus to a 2000-2017 timeframe [Angrist et al., 2021]. Our approach here sacrifices some of the richness of the learning assessments used in this test linking literature to allow us to (a) avoid difficult questions about harmonization by relying on identical survey questions across countries, and (b) directly measure literacy across cohorts spanning multiple decades without imputation or extrapolation.

[5] While statistics for women are nationally representative, those for men are often not perfectly so, because men are sampled only from households with adult women present.

The surveys do not test individuals with secondary schooling and, in a few MICS surveys, also lack literacy data on individuals with no schooling. In the latter case, we make the assumption that individuals with no schooling are illiterate.[6] We follow the DHS and MICS convention of assuming that individuals with secondary education are literate to construct trends of literacy.[7]

Note this missing data problem has little effect on our analysis. While the unconditional literacy rates reported below rest on the assumptions we make about literacy at the top and bottom of the schooling distribution, the school quality measures that are the focus of the paper are much less fragile to this missing data problem, as they focus on individuals with five years of primary schooling, all of whom are tested.

Individuals chose the language they want for the literacy test. In most countries, reading cards have been translated into all local languages and across the whole woman sample there are only 0.25 percent of tests that could not take place because of the lack of availability of reading cards in the local language. In a small share of interviews (0.17 percent in the woman sample), individuals could not be tested because they are visually impaired. These observations are excluded from the analysis and the final woman sample includes 99.5 percent of the observations.

Overall, data coverage for men is lower than for women, with fewer countries covered and a total sample size five times smaller (4.6 million observations for women and 1 million for men). Samples of women are nationally representative, although in seven countries only ever-married women have been interviewed [8] but samples for men are not perfectly representative, as men are selected only from households with women respondents, excluding men living alone.

All available DHS and MICS datasets for a country are used for our estimates, with the exception of pre-2000 surveys that collected self-reported literacy. To the exception of

[6]Where these observations are not censored, the median literacy rates for individuals with zero schooling are 2 percent for women and 5 percent for men, based on all surveys with at least 10 percent of the population reporting zero schooling. In only 34 surveys out of 242 for women and 22 of 93 for men, the rate of literacy is higher than 10 percent for individuals with no schooling.

[7]This assumption is questionable in some countries such as Ghana where literacy rates at the end of primary school are still low, and a few recent DHS surveys that test all adults allow us to test (and reject) this assumption. The numbers of surveys with full coverage is too low, and too recent, to serve as the basis of our main analysis.

[8]Only ever-married women were interviewed in Afghanistan, Bangladesh, Egypt, Indonesia, Jordan, Nepal and Pakistan. Given that most women eventually marry in these countries, potential selection biases should be minimal.

Panama and Trinidad-and-Tobago, no country in our dataset is classified as high-income countries,[9] and so our estimates should be seen as representative of non high-income countries. Data coverage is excellent in the woman and man samples for Sub-Saharan Africa and South Asia[10] with 99 percent of the population covered for women and men in South Asia and 97 and 86 percent of women and men respectively in Sub-Saharan Africa (see table 1). Coverage is also good for Middle East and North Africa for women (75 percent of the population excluding high income countries) but in this region men are not surveyed and there no available survey. Similarly, in Latin America and Caribbean the coverage is much better for women (41 percent) than for men (10 percent) and the coverage of the region suffers from the absence of data for Brazil, whose surveys pre-date 2000 and only includes self-reported literacy measures, and the fact that there is only one survey available for Mexico. Because there is no available data for China, the coverage for the East and Asia Pacific region is 30 percent for women and 26 percent for men but goes up to and 91 and 77 percent for women and men respectively when China is excluded. Finally, there are only few countries with available data in the European and Central Asia region making up 5 and 1 percent of the population of the region for women and men respectively. In this region, some countries are excluded (e.g. Tajikistan, Ukraine) because virtually all individuals reach secondary school making it impossible to estimate our measure of school quality on individuals with primary schooling or less. Because of the low coverage in the Europe and Central Asia region, we do not report regional averages for this region. Similarly, no regional averages are reported for men for the Middle East and North Africa and the Latin America and Caribbean regions.

There are a large number of individuals in the samples who are still in primary school between the age of 15 and 19. 21 (17) percent of women (men) are still in primary school at age 15 and rates decrease to 2 and 1 percent for women and men respectively by age 20. Including individuals who are still in primary school could bias our estimates as the composition of cohorts might change between two surveys. For this reason, samples are restricted to individuals aged 20 to 49.

DHS and MICS surveys provide information on the number of years of education an individual has reached. We only consider formal education in our estimates. In the few cases where the length of primary schooling has changed between two surveys (e.g. Angola), we set the length of primary to the lowest value of the two surveys and considered that

[9]World Bank classification as of 1st of July 2020 https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

[10]Region classification follows World Bank classification. See here for the complete list: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

individuals who exceeded this value in the other survey have reached secondary school and are literate. Age is computed as the mean year of interview minus year of birth in all surveys. We do so to avoid any variance in age for individuals born the same year and interviewed during the same survey. Indeed, in some cases, surveys take a long time to administer and places surveyed first may have different characteristics than later-surveyed places, which might create some spurious age effects.

3 Measuring education quality

Although measuring the *quantity* of education of a population is usually straightforward using administrative data or household surveys, measuring the *quality* of this education remains challenging. Large national or international assessment programs make considerable efforts to measure learning in school in a reliable way via standardized tests of representative student samples. However, because of the cost and complexity of collecting such data, standardized tests happen only infrequently, and very few low- and lower-middle income countries participate in internationally comparable assessments whatsoever. In order to measure educational quality in a large number of developing countries and over time, we propose a measure of educational quality based on data available in household surveys, following Oye et al. [2016], and Gollin et al. [2014]. We use the adult literacy rate as our proxy measure of human capital, and conceive of the quantity and quality of schooling as inputs in the literacy production function. The main advantage of literacy as an outcome is that it is available across a wide range of countries and time periods and is measured in a comparable way. While literacy alone is clearly a low bar against which to judge education systems, ceiling effects are less worrisome in our context, as only a handful of countries report rates above 90 percent in any year.

We define educational quality as the tendency for schooling to generate learning. More precisely, we measure quality as the expected level of literacy conditional on five years of schooling.[11]. Primary schooling lasts between four to eight years in sampled countries, with six years being the most common length (63 percent of countries in our sample). Five years is thus close to the end of primary in most countries.

[11]This measure, essentially capturing the strength of the cross-sectional correlation between schooling and literacy serves as a substitute for a more ideal measure, where longitudinal panel data is available, tracking learning as students progress through grades. See Muralidharan et al. [2019], Andrabi et al. [2011], Singh [2019]

More formally, let $\mathsf{L}_{ijt} \in \{0, 1, 2\}$ be the level of literacy and G_{ijt} the maximum grade attained of an individual i in country j at time t. Our indicator of educational quality faced by an individual in j born in year t is defined as the predicted value $\hat{\mathsf{L}_{jt}}$ at $\mathsf{G} = 5$ from the following regression:

$$\mathsf{L}_{ijt} = \alpha_{jt} + \beta_{jt}\mathsf{G}_{ijt} + u_{ijt} \tag{1}$$

Obvious alternative definitions include using the estimated $\hat{\beta}$ or the simple average of literacy conditional on five years of schooling. The former has a weaker correlation with independent measures of school quality discussed in the next section. The latter performs quite similarly to our preferred measure. One advantage of our regression-based approach is that it allows us to make use of the full dataset, whereas the number of individuals with exactly five years of schooling can become small for some country-sex-birth year cells.

3.1 Comparison with international test scores

To assess the validity of our measure, we briefly examine its cross-country correlation with alternative measures of school quality, as well as per capita GDP and education spending, before turning to our primary analysis of trends over time.

Across countries, our measure of average education quality shows a significant and roughly linear, positive relationship with independent measures of student performance based on more detailed standardized tests – though the correspondence is far from perfect.

No single standardized test covers a large share of developing countries. To examine correlations across a relatively large number of countries, we draw on results reported by Altinok et al. [2018] who provide rough correspondences to link scores across international and regional tests on a common scale. The resulting scores, produced by the World Bank under the rubric of Harmonized Learning Outcomes (HLO), combine national averages from various underlying assessments (PASEC, LLECE, EGRA, PISA, SACMEQ and TIMSS/PIRLS) spanning 163 countries. As time-series variation is limited, we focus here on cross-country patterns at a point in time.

We match the birth cohorts in our data to the corresponding survey year and estimate age for each of the tests underlying HLO estimates. Based on the grade at which the tests are administered, we assume the average age of test takers for PASEC, SACMECQ and LLECE is 12 years, for EGRA is 8 years, 10 for TIMSS/PIRLS, and 15 for PISA. We can match 117 country, sex and survey-year cells from our data to the HLO estimates, spanning 40 countries.

Figure 1a shows the comparison between our measure of expected literacy at grade five and HLO estimates. We observe a relatively linear relationship, with one standard deviation in HLO (100 points) equivalent to about 18 percentage points of expected literacy at grade five. The Spearman correlation is 0.41 for the whole sample with no meaningful difference between men and women.

3.2 Comparison with income per capita and educational spending

We also see that richer countries and countries that spend more on education tend to have better school quality by our measure, though again there are some notable exceptions. The estimated correlation coefficients between school quality and GDP per capita or education spending, respectively, are high and positive (0.4 and 0.5). Figure 1b also shows that a group of countries perform well despite low GDP per capita and education spending such as Burundi, Rwanda, Madagascar and Lesotho in Africa; Myanmar, Nepal, and Vietnam in Asia.

4 Empirical strategy for disentangling age, period, and cohort effects

The primary goal of the empirical analysis that follows is to measure the evolution of the quality of schooling over time — i.e., trends in literacy conditional on years of schooling. Recall that our survey data reports literacy and schooling for repeated cross-sections of women and men of various ages, observed at multiple survey years in each country. Thus our core challenge is to disentangle the role of age, period, and cohort effects in these repeated cross-sections.

Age-period-cohort (APC) models of this sort are widely used in demography, sociology, political science and, to a lesser degree, economics (see Fosse and Winship [2019] for a recent review). In our setting, age captures life-cycle effects on literacy, which may evolve even after an individual leaves school if they acquire human capital throughout their career, or instead it decays as time passes. Period effects allow for the possibility that contemporaneous events may affect measured literacy, e.g., if economic booms increase formal employment which contributes to literacy through on-the-job learning. Cohort effects, which are our primary object of interest, capture differences in literacy conditional on schooling for individuals born (and hence educated) in different years. These cohort effects capture secular trends in schools' value added, as well as trends in other non-school inputs into learning such as early-childhood nutrition and – as discussed at length below – changes in selection into schooling based on scholastic ability.

The basic model poses a fundamental identification challenge. If we posit that age, period, and cohort effects are linear and additive, there is an obvious, perfect collinearity between the three factors (age = period − cohort), making it impossible to estimate the model without invoking additional functional-form assumptions. The following sub-sections briefly motivate the modeling assumptions we invoke, and our reasons for eschewing alternative approaches.

4.1 Assumptions underlying our modeling approach

In most of what follows we adopt one of the most common approaches to estimating APC models in the economic literature: the Deaton-Paxson (DP) normalization [Deaton and Paxson, 1994].[12] The DP approach imposes a parametric restriction on period effects: specifically, that the period effects do not contain any linear time trend. In practical terms, the variables are detrended and the period effects are restricted to be orthogonal to a linear time trend. This approach provides a solution to the APC identification problem and the model is just identified. OLS regressions or other estimation methods can then be used to estimate coefficients for age, period, and cohort. However, the DP method requires a minimum of three surveys for the model to be identified. A shorter cohort cycle can yield biased estimates as the method would attribute any variation in literacy between the first and second survey to age and cohort effects.

The choice of the best method to estimate our APC model depends on what assumptions we're willing to make about the true age, period and cohort effects. First, we do not want to impose restrictions on *cohort effects*, whose estimation is our main research goal.

Second, we want to allow that there might be *age effects*, although we do not have a strong prior on their direction. The estimation of age effects is also a secondary research goal for us. Finding evidence for negative age effects would demonstrate that some of the

[12] See also Hanoch and Honig [1985] for an earlier presentation of a similar approach.

education is wasted but finding positive age effects would show that the consequences of low quality education can be overcome as individuals age. Because of the limited number of large longitudinal datasets including information on adult cognitive skills, the process of literacy retention or acquisition after students leave school in developing countries is still not very well known. There is some limited evidence that literacy skills continue to evolve after individuals leave school and that changes in adult literacy vary by sex, grate attained or personal circumstances. For instance, Gorman and Pollitt [1997], using a longitudinal dataset of teenagers in Guatemala, find that literacy continues to increase after students leave school but that the increase was stronger for students who completed fewer grades. In a longitudinal study of teenagers in Malawi, Soler-Hampejsek et al. [2018] find that literacy levels in English tend to deteriorate after students leave school but skills in local language remain the same. Moreover, they find that grade attained, rather than life experience or personal characteristics, is the main determinant of variability in literacy retention, such as the ones who achieved more grades were more able to keep their level of literacy after leaving school. They also find that girls tend to lose literacy skills faster than boys. In another study using the same data from Malawi and longitudinal data from Zambia and Bangladesh, Psaki et al. [2019] find that early child-bearing is associated with larger losses in literacy skills for adolescent girls, with stronger effects for girls with low attainment (grade seven or below). The three previous studies are focused on learning retention of adolescents but, to our knowledge, there is no longitudinal study on literacy changes for adults. It has been found that adults who take part in adult literacy programs often relapse to illiteracy, which might be due to low proficiency levels, lack of use of literacy skills or difficulty in teaching people how to read at an age where the brain is less malleable [Abadzi, 2003].

Third, to achieve identification we assume that *period effects* do not exhibit a trend, that is there is no societal change that allows adults to learn (or forget) how to read. The roll out of adult literacy programs could be a factor of a positive trend in period effects (or negative if they tend to be closed down) but their effectiveness has been questioned and they often do not reach a large share of the population. Adoption of mobile phones could also be a factor in a positive trend in period effects but the relative short span of our data (on average 10 years and maximum 20 years) may not be sufficient to see a change. Note that we allow for the possibility of systematic differences between different periods because of survey effects. Indeed, although DHS and MICS surveys are of good quality, challenges in collecting data in developing countries might lead to some sampling variance between surveys and changes in reading cards or interviewer training could slightly change the grading of on the literacy scale. However, we see no reason to anticipate a systematic trend in these survey effects across multiple rounds. For these reasons, using a DP normalization, which assumes no period trends, seem the most appropriate but we will first turn a visual description of our data to see how our assumptions hold.

4.2 The Deaton-Paxson normalization

The DP normalization can be written as:

$$\mathsf{L}_{iban} = \sum_{b=min+1}^{max} (\alpha_b + \beta_b \mathsf{G}_{ib})\mathsf{C}_b + \sum_{a=16}^{49} (\delta_a \mathsf{A}_a) + \sum_{n=3}^{N} (\theta_n \mathsf{S}_n^*) + u_{iban} \tag{2}$$

where a is the age of an individual interviewed in survey n and born in cohort b. C_b denotes cohort fixed effects, and G_{ib} is the number of completed years of primary schooling. A_a takes the value one if an individual is aged a. Age fixed-effects estimated in model 2 reflect systematic differences in literacy rates for individuals born the same year but interviewed at different ages. S_n^* is a set of $N-2$ period dummies that have been transformed to be make period effects effects orthogonal to a time-trend and add to zero. Adapting the notation of Deaton [1997], the period dummies are transformed such that:

$$\mathsf{S}_n^* = \mathsf{S}_n - [(n-1)S_2 - (n-2)S_1] \tag{3}$$

where S_n are the raw period dummies.

We perform estimates separately by sex for each country with at least three available surveys and 500 observations and include individuals with some primary schooling. The model includes fixed effects for age, cohort and period (survey). [13]

The large number of fixed effects can make models difficult to estimate, especially when using a Tobit estimator. We also test a more parsimonious model where age fixed effects are replaced by age and age squared as described in equation (4).

[13] We omit individuals with zero years of schooling as age effects for them may differ and their inclusion could bias our predictions. Individuals with no schooling have on average very low levels of literacy (5.5 percent and 9.5 percent for 20 year-old women and men respectively) and so it can be expected that age effects for them will be non-negative. An APC decomposition of literacy using the DP normalization for people with no schooling shows small and positive age effects (median effect: 0.08 pp per year or 2.4 pp over 30 years) and very small cohort effects (median effect 0.03 pp per year). See figure 10 in Appendix. These samll effects imply people with no schooling usually do not manage to learn how to read later in life.

$$\mathsf{L}_{iban} = \sum_{b=min+1}^{max} (\alpha_b + \beta_b \mathsf{G}_{ib})\mathsf{C}_b + \gamma \mathsf{A}_a + \delta \mathsf{A}_a^2 + \sum_{n=3}^{N} (\theta_n \mathsf{S}_n) + u_{iban} \qquad (4)$$

We estimate equations (2) and (4) for 24 countries for men and 47 for women. Literacy rates are bounded in the interval [0;1] and the Tobit estimator avoids predictions outside this interval which, in principle, may influence the magnitude of age and cohort effects. In practice, a summary of the trends in age and cohort effects in Figure 11 in the appendix shows that OLS and Tobit estimates give similar results.

4.3 A note on countries with only one or two survey rounds

As mentioned earlier, there are several cases of country-sex pairs with only two available surveys (38 cases) or even one (36 cases), rendering the DP normalization impossible. When two surveys are available, it is possible to fit an age-cohort model (AC) ignoring the effect of period or survey. In essence, the model attributes the entire difference in literacy between surveys for the same cohort to age effects, which given our assumptions of zero or small period effects seems reasonable. The risk is that if the differences between two surveys actually reflects a survey effect, spurious age effects could be estimated, which would bias cohort effects. In order to estimate an AC model, only controls for age and cohort effects are included.. The model is estimated by including age and age squared instead of age fixed effects. In the case of only one available survey, the fitted model is a pure cohort model (C), as no separate age and period effects can be estimated. Since, in most cases, age effects tend to be negative not controlling for it could bias cohort trends upward.

To test the sensitivity of our results to the inclusion of age and period effects, we run an AC and C models on country/sex cases with at least three available surveys and compare it to the APC results. Results can be seen in figure 12 in Appendix. Ignoring period effects does not change the estimated age and cohort effects in a meaningful way and makes it possible to obtain reliable estimates when only two surveys are available. However, the cohort-only model produces trends in cohort effects which tend to be flatter. Nevertheless, we report results from countries with only one available survey using this cohort-only approach, as the trends across recent cohorts remain informative, but it seems likely that the long term-trends here may underestimate the true decline in school quality.

5 Main results: cohort effects in education quality around the world

Estimating the age-period-cohort model described above provides us with measures of cohort effects in education quality for each country in our sample: 87 countries in the case of women, and 58 for men. In most countries the trends span about 40 years, from the cohort born in the mid 1950's to the cohort born in the mid 1990's. All the results discussed in this section are based on the Tobit model with age and age squared for countries presented above, but the qualitative conclusions of our discussion are not driven by the choice of the specification or by the inclusion of cases with one or two available surveys.

5.1 Graphical analysis for selected large countries

Following Yang and Land (2013), we first present a visual description of the data to detect potential age or period effects and see if our assumptions on age and period effects might hold. Figure 2 shows the literacy rate across cohorts and surveys for women with four to six years of primary schooling in the eight largest countries in our dataset with at least four surveys, with the exception of India where we opt to use only two of the extant survey rounds for reasons detailed in the appendix.[14] If there were no age, period or survey effects, we would expect that literacy rates for women born the same year but interviewed at different dates would be on average the same and literacy rate trends estimated with different surveys would overlap. This tends to be the case in countries such as Indonesia, Nigeria or Egypt. In these three countries, it is unlikely that there are large period or age effects as women born the same year tend to have roughly the same literacy rate when they are interviewed at different dates.[15] In the cases of Bangladesh, Tanzania or Kenya, women in the same birth cohort show fairly large differences in literacy across survey rounds, but we see relatively few differences in literacy across birth cohorts in the same survey round, suggesting that differences might be due to survey effects. Finally, in the case of Ethiopia and India, there is a systematic pattern of women interviewed in more recent surveys showing lower rates of literacy, suggesting negative period or age effects. This seems particularly large in Ethiopia, where women born the same year lost about 30 percentage points of literacy between the

[14] India alone accounts for about one third of the total population of the countries in our sample.

[15] Without additional assumptions, of course, we cannot rule out the possibility that large age and period effects of opposite sign cancel each other out.

2000 and 2016 survey. The 10 percentage-point loss in India in 10 years is also quite large, especially because the two most recent rounds show no age decay. After careful scrutiny of the India data (see Appendix A.4), we conclude that the first round of Indian data (NFHS-3) is not comparable and should be excluded from analysis.

Systematic differences in literacy rates between surveys could also be due to differences in sampling. If this were the case, we would also expect to see differences across survey rounds in other indicators, including time invariant characteristics like schooling. However, we show in figure 3 that the share of women born in the same cohort accessing grade 5 is very stable across surveys. For instance, while literacy conditional on schooling for women in a given cohort in Ethiopia appears to decline between survey rounds systematically from 2000 and 2019, schooling does not. This pattern is consistent with the idea that literacy may decay over the lifecycle, but schooling, by definition, cannot.

We turn now from 'raw' data to the core estimation results. Figure 4 shows the estimated age, period and cohort effects for women for the eight large countries discussed previously. The estimates are based on the Tobit estimator using the Deaton-Paxson normalization. In all countries, period effects tend to be very small, suggesting that there are no large systematic differences in literacy rates between survey rounds once we control for age and cohort effects.

Age effects, shown in the first column of Figure 4 tend to be either flat or declining. This suggests that for this subset of countries – especially Bangladesh and Ethiopia – rather than acquiring human capital in the form of literacy over their life cycle, women tend to lose it through a process of decay during their adult lifetimes. Kenya and Nigeria also show slight evidence of such decay, with only Egypt showing any sign of human capital acquisition over women's adult lifetimes.

The trends in cohort effects, which are our main object of interest, show a fairly sharp downward trend in Nigeria and Ethiopia. Taken at face value, the estimates for Ethiopia suggest the probability a woman with five years of schooling would become literate was nearly 100 percent for the 1960s birth cohorts, but had fallen to roughly 60 percent for the mid-1990s cohorts. Nigeria started off at a lower baseline level of education quality, and has reached a lower point today: with only about 20 percent of adult women educated in the late 1990s able to read.

The other five countries in Figure 4 show relatively flat trends in education quality, as measured by cohort effects. Their quality levels though, are far from uniform: while five years of schooling are a virtual guarantee of literacy in Indonesia, that is true for only less than half of educated women in Egypt. The remaining large countries (Bangladesh, Kenya, and Tanzania) fall somewhere in between, and all of them outperform Nigeria.

5.2 On average, education quality in the developing world has declined steadily over multiple decades, albeit with significant heterogeneity in levels and trends

Similar patterns hold for a broader set of countries. For brevity we focus on regional averages. These are a simple average of the countries in the sample, without population weights.

We focus on three main indicators to document changes in educational outcomes: share of birth cohort accessing grade five, as well as the literacy rate and average education quality (i.e., estimated literacy for an adult with five years of schooling) where the latter two statistics are both corrected for age and period effects per the Deaton-Paxson normalization described above.

Several broad patterns stand out in the regional trends shown in Tables 2 and 3.

First, most of the gains in literacy are associated with school expansion (rather than any improvement in education quality, as we show below). Schooling figures show consistent progress in access to education for both men and women across all regions. From the cohort born in the 1950's to the cohort born in the 1990's, women's access to five or more years of schooling has increased by 60 percentage points in South Asia, 40 percentage points in Sub-Saharan Africa, and reached above 90 percent in East Asia and Pacific and Latin America regions. Improvements for men were slightly smaller as they started from higher levels but were still large (+15 pp in East Asia, +50 pp in South Asia and +20 pp in Sub-Saharan Africa). These large gains were accompanied by similar increases in literacy rates across cohorts.

Second, there is large variance across countries in the level of education quality. In Ghana, Guinea, the Gambia, Sierra Leone and Nigeria 20 percent or less of women are able to read at grade five whereas in Vietnam, Rwanda, Burundi, El Salvador, Guatemala, Honduras, Costa Rica and Bolivia more than 95 percent can read by the end of grade five.

Third, despite big cross-country differences in education quality, there are very few cases of dramatic improvements in education quality. In most cases, cross-country gaps in education quality seem to have emerged at least several decades ago. Trends over time are generally flat or declining. In over two thirds of countries in our sample, the difference between the education quality faced by the oldest and youngest decadal cohorts of women is less than 10 percentage points. However, where changes in education quality occure, they tend to be negative. For 38 of 87 countries in our sample of women, education quality did not change by more than 10 percentage points between the cohort born in the 1960's and the cohort born in the 1990's, while it fell by more than this margin in 46 countries and rose in just 2 (see Figure 6). [16]

Finally, despite the general picture of stagnation, regional trends show strikingly different patterns in the evolution of school access and – the more novel contribution of this paper – education quality. To highlight the latter point, Figure 5 shows the trends in raw, unconditional literacy and education quality (i.e., literacy conditional on schooling) for women and men on a consistent sample of countries from the birth cohort born in 1955 to the birth cohort born in 1998. We observe five regions that have followed different trends in educational outcomes.

East Asia and Latin America show steady, gradual increases in literacy, and relative stability in education quality over multiple decades. The two regions with the largest gains in literacy are the Middle East and North Africa and South Asia. The former has seen fairly stagnant education quality, while South Asia has seen a moderate decline. Finally, Sub-Saharan Africa has experienced steady progress in literacy, especially for women; however, it has experienced the largest drop in education quality (a fall of roughly 25 pp between the '60s and '90s cohorts for both women and men).

5.3 The gender gap in education quality has decreased over time (in conjunction with a declining gender gap in school access)

Education quality for men and women can be compared for a sub-sample of countries and cohorts (see Table 4). In Sub-Saharan Africa, the number of countries with a positive gender

[16] The time series for the 87th country, Thailand, is missing for the 1960s. Additionally, we should note that in one of the best-performing countries, Vietnam, it is impossible to construct our measure of education quality past the mid-1990s. There are no longer any survey respondents with (only) 4 to 6 years of schooling.

gap in favor of men has remained stable between the 1960's to the 1990's with 12 countries out of 35 countries with data for men and women. The gender gap in education quality has decreased slightly but remained very high (+9 percentage points for men). Unlike Sub-Saharan Africa, the gender gap has decreased in South Asia and education quality for men and women has converged over the past three decades and education quality for the last coohort born in the 1990's is almost identical. In East Asia and Pacific, the gender gap has remained positive for women between the 1960s and the 1980s.

5.4 Education quality is correlated with long-term growth

Our estimates allow us to revisit the relationship between schooling, education quality and economic growth [Mankiw et al., 1992, Wossmann and Hanushek, 2007]. Because our estimates of education quality and schooling cover longer time-spans, we are able to estimate the stock of human capital available at different dates and see if it explains subsequent economic growth. We follow standard practice in the literature and specify the growth equation as a function of initial income per capita and the stock of human capital measured by the quantity of schooling and education quality. We use our measure of education quality, expected literacy at grade five, as a measure of education quality and two separate measures of school quantity: the number of years of schooling of the adult population following Barro and Lee [2013], and the share of the population accessing grade five from our own estimates, such that:

$$Y_{i,2019} - Y_{it} = \alpha + \beta Y_{it} + \gamma S_{it} + \delta Q_{it} + u_{it} \tag{5}$$

where Y_{it} is log per capita GDP in country i in year t, growth on the left-hand side is measure up to the latest available data (2019), S_{it} is the quantity of schooling of the adult population, and Q_{it} is the average expected literacy rate at grade five of the population aged 20 to 40. Note that both education quantity and quality are measured in the baseline year. Quantity is either measured as the number of years of schooling of the adult population aged 20 to 40 from Barro and Lee [2013], or as the share of the female population aged 20 to 40 (or the highest available age) with at least five years of schooling in year i.

Results are presented in Table 5. Using either definition of education quantity, there is a correlation between initial education quality and subsequent economic growth. For instance, in column (5), economic growth from 2010 to 2019 was 0.7 percentage points higher in countries with one standard deviation (sd=0.23) higher education quality. These associations between education quality and growth appear, if anything, to be increasing in magnitude over time.

6 Accounting for changing patterns of selection into school

So far we have followed the convention in the education assessment literature by defining 'school quality' as a simple conditional average, the literacy rate conditional on five years of schooling. Obviously, the evolution of school quality defined this way may reflect changes in schools' value added, or changes in the composition of pupils. The expansion of schooling under free primary education reforms in the 1990s and 2000s, for instance, incorporated a growing share of children from more disadvantaged backgrounds who might have more difficulty learning to read [Deininger, 2003, Bold et al., 2015].

In a study using panel data from Vietnam, Peru, India and Ethiopia, Singh [2019] finds that differences in test scores between countries are mainly due to differences in school quality rather than differences in child's endowments, suggesting that differences in composition of students may not be driving our results. The literature on the Mincerian wage returns to schooling is also potentially informative here. While we are not concerned with earnings, there is a general consensus that OLS estimate of the returns to schooling, while theoretically subject to the same upward bias due to unobserved ability that concerns us here, appear quite similar to well-identified estimates from natural experiments Card [1999]. Nevertheless, there is recent evidence from secondary schooling in Ghana that (upward) selection bias is sometimes real and significant [Ware, E. O., Kornu, D. D. K., & Adusei, 2018].

In the absence of a valid instrumental variable for access to schooling, we do not attempt to provide a point estimate of the impact of selection on school quality. Instead, we attempt to provide some insight into the possible magnitude of selection and how the cohorts of women and men have changed over time.

6.1 Comparing height between women who did and did not go to school as an indicator of selection

Ideally, we could compare the academic ability and socioeconomic status during childhood of individuals who drop out of school at diffferent points, to examine patterns of selection directly. This is not feasible using our data based on surveys of adult women and men. However, adult height provides a rough but well-established indicator of childhood health and nutrition [Garenne, 2011, Currie and Vogl, 2013]. Adult height is recorded for women in 64 countries in our dataset. We focus on women here because of limited anthropometric data coverage for men. Figure 7 shows the trends in female adult heights by schooling status.

Trends in adult height shed light on two ways non-school inputs in the human capital production function may bias estimates of trends in school quality. First, long-term improvements in overall child well-being could contribute to improvements in cognitive skills, creating a spurious impression of improved school productivity. Overall the height of women with at least some schooling has risen over time for a slim majority (37 of 64) of countries in our sample, with an average rate of increase under 0.1 centimeters per year.

Second, changes in the height-gap between women with and without schooling may tell us something about the degree of selection into schooling over time. Looking at all women (rather than just those with schooling), we see even stronger positive trends in heightt, with a positive increase in 44 of 64 countries and an average increase of about 0.2 centimeters per year. In short, the height-advantage of girls in school appears to have narrowed over time. This could be due to an improvement in nutrition at the bottom of the distribution or, alternatively, a reduction in the degree of selection into school on socioeconomic status. The latter alternative would be consistent with the idea that the expansion of school systems incorporated relatively more disadvantaged students. Indeed for women born in the 1960's, the average adult height of thos who went to school was 0.47 centimeters taller than the general population, but that difference shrunk to 0.27 centimeter for women born in 1990's.

6.2 Does measured school quality fall when access to schooling expands?

The correlation between the share of individuals accessing school and school quality is positive across the whole sample (r=0.23 when pooling men and women). This positive relationship is mainly driven by the correlation between countries (r=0.43 for men and r=0.38 for women), while the correlations within countries are negative (r=-0.3 for men and r=-0.38 women). Countries with higher access to schooling have better school quality, but as the share of individuals accessing school increases within a country, school performance decreases.

However, these correlations could be biased by time trends in education expansion and school quality. The association between selection into school and school quality can be tested more formally using results from cohort estimates that provide long time series for each country. Table 6 shows the results of country fixed-effects models for women and men respectively where the dependent variable is school quality from our APC model above. We control for year of birth to allow for trends in underlying share quality, and focus on the coefficient on the share of a cohort with some schooling, as well as the share with secondary schooling, to capture the role of selection into school. We also include administrative data on repetition rates, available from the 1970's onward, to test if changes in school quality can be explained by changes in repetition practices[17] Finally, for a sub-sample of countries, we control for the height of women with some schooling.

Column (1) shows downward trends of about -4.6 and -5.9 percentage points per decade in school quality. for women and men, respectively. Controlling for the share of the cohort that has attended school or has reached secondary schooling (column (2)) has little impact in the men's sample but reduces the magnitude of the decline among women by almost 2 points. Of course, this attenuation of the coefficient on birth year could be due changes in selection of students or to the fact that as school systems expand it is harder and harder for governments to keep quality constant.

Controlling for repetition rates in column (3) has little effect on the estimated secular trend. However, repetition rates may be of independent interest, and are strongly associated with changes in school quality for both men and women. An extra year of schooling due to repetition is associated with a 12.1 and 6.8 percentage points increase in school quality for women and men, respectively. This is a relatively large effect; our estimates show that about 73 percent of men and women can read after five years of schooling, equivalent to about 14.6 percentage points of extra literacy per year of schooling. Our coefficient on repetition would suggest that an extra year due to repetition is worth about half to more than three quarters of a year of regular schooling. Notably, some of the best performing countries have very high levels of repetition (e.g. Madagascar, Rwanda or Burundi).

Finally, in column (5), we control for the height of women with schooling, which is positive

[17] Repetition data is taken from the UNESCO Instiute for Staitstics. Data are matched for men and women separately. Gaps in data are interpolated. We match repetition rates to cohorts when students are 10 years old. The indicator for repetition rate is the cumulative repetition rates from grade one to five, that is the number of extra years of schooling a student would have at grade five if she had experienced her country repetition rates when she was 10.

and statistically significant at 5 percent. Once again, controlling for this association does little to change the trend in measured school quality.

Overall, selection into school and changes in repetition practices may explain some of the downward trend in school quality but, in a simple regresion framework, fail to account for all of it.

6.3 Free primary education as an exogenous shift in access

Over the past 50 years, many developing countries abolished user fees for primary school. In at least some cases, these free primary education (FPE) reforms led to a significant increase in schooling levels at the national level [Deininger, 2003, Osili and Long, 2008, Bold et al., 2015].These putatively exogenous, policy-induced increases in access to schooling can shed light on how enrollment changes affected trends in school quality.

FPE policies could affect school quality through several channels. First,lowering barriers to school entry might change the composition of pupils. Second, FPE clearly increased the number of pupils in some countries which, absent comparable increases in staffing and school expenditure, would lead to a deterioriation in resources per pupil. Third, while more speculative FPE may have undermined the accountability of school officials to local communities..In short, while FPE does not isolate the role of selection into schooling, it does offer the possibility of overcoming reciprocal causation between school quality and enrollment.

We examine 51 FPE reforms across 27 developing countries compiled by Crawfurd and Ali [2022], drawing on earlier compilations from Bhalotra et al. [2014] and Tomasevski [2006], among others, for which we have overlapping data. For each reform episode, we estimate an interrupted time series model with a break for the birth cohort that was eight years-old when fees were abolished.[18] We restrict analysis to countries with data for cohorts born at least 10 years on either side of the policy introduction, and present estimates for up to 20 years on either side. Estimates were weighted by the size of the sample of the cohort to give more weight to more precise estimates.

The interrupted time series estimates of the share of students accessing grade five before and after the FPE reforms show that the mean trend increased from 0.7 percentage points

[18] Results are robust to alternative age cutoffs at 6, 8, 10 and 12 years old when the policy was introduced.

per year to 1.2 percentage points per year after fees were abolished. Figure 8 illustrates these calculations for three cases with particularly large accelerations in schooling after the abolition of fees: women in Ethiopia, Madagascar, and Mali. In each case the acceleration in the increase in average schooling was accompanied by an acceleration in the decline in school quality (or, in the case of Mali, a sudden drop in quality and continued decline).

A summary of these changes in trends in school quality before and after FPE reforms is shown for all available countries in Figure 9 for women and men separately. In 27 cases, we observe no statistically significant change in the trend of school quality before and after the implementation of the FPE reform. There are more statistically significant negative effects than positive ones (20 vs 3) and the mean effects are -0.4 percentage point per year for women and -0.9 for men with only few cases of a very rapid drop in school quality after the introduction of FPE reforms. Although the general trends in school quality were already decreasing before the FPE reforms (-0.2 percentage points for women and men), average declines in school quality have accelerated substantially after the reforms (-0.6 and -1.1 percentage points per year for women and men respectively) and could explain a large share of the drop in quality in the sub-sample of countries having introduced FPE.

Overall, the data suggest some role for selection effects in depressing school quality, but several signs suggest selection alone cannot explain the long-term trends we observe. Various patterns point in opposite directions. For instance, by one measure the composition of students has improved over time: students in later decades appear healthier than their peers in earlier decades, as proxied by their adult height. But selection into schooling on (future) height has declined. Looking across countries and over time, a larger share of students accessing schooling is associated with a drop in measured quality. But controlling for enrollment changes in a simple regression framework fails to explain away the long-term deterioration in school quality overall. Similarly, FPE reforms are associated with an acceleration both schooling levels and of the negative trend in school quality. But school quality was decreasing before the introduction of these reforms and FPE reforms might also have been associated with a reduction of resources per students.

7 Conclusion

We use literacy tests in survey data to construct long-term trends in literacy for 87 developing countries, spanning birth cohorts from the 1950s to 2000. We show that over this period literacy rates have increased substantially, but virtually all progress has been due to the increase in access to school rather than any improvement school quality, measured as the propensity for schooling to generate literacy after five years of schooling.

Overall, our measure of school quality is low in developing countries, with about 70% of women able to read after grade five, and has been declining over time. From the 1950s to 2000, school quality deterioration implied the probability that a woman with five years of schooling could read a sentence fell by roughly two to four percentage points per decade, and about six percentage points for our smaller sample of countries with data on men. Nevertheless, literacy gains associated with increased schooling rates far outpaced the offsetting effect of declining quality, i.e., literacy conditional on schooling.

Two limitations of our approach bear repeating. First, our reliance on simple literacy tests as a measure of human capital clearly ignores important higher-order skills, not to mention numeracy and other domains. Nevertheless, we argue that literacy remains a useful proxy for learning levels in many school systems, where literacy is far from universal, and puts our focus on the lower parts of the learning distribution which are often ignored in favor of policy attention to students passing competitive exams for elite secondary and tertiary schools.

A second limitation is our inability to precisely quantify the role of changing patterns of selection into schooling (or into higher grades) in explaining the changes in school quality that we observe. Of particular concern is that reduced selectivity, e.g., enrollment of more low-income or otherwise under-prepared students as schooling systems have expanded, might create a spurious impression of declining school quality. We have shown that countries and periods with faster enrollment growth did indeed experience significantly faster deterioration in school quality. But allowing for these effects does not appear to overturn our main finding of declining school quality. Furthermore, correcting for differences across countries and periods in grade repetition rates, or the height of individuals (an indicator of socio-economic status that is mostly determined prior to schooling) also does little to change the picture of declining quality.

From a policy perspective, the fact that most developing countries with high (unconditional) literacy rates achieved them primarily by increasing average years of schooling rather than school quality may be instructive for lagging countries. Expanding access to schooling has produced remarkable gains in overall literacy, while not much else has.

Turning to implications for future work, our hope in embarking on this project was that by constructing long-term trends in school quality, we could identify countries which have made transformational changes, and risen quickly in global league tables of learning outcomes. Such exemplars appear not to exist. Instead, the rank-ordering of countries' school quality has been relatively stable over decades. Nevertheless, moderate improvements over the period of a decade or more do exist, and future research could usefully complement the large literature on policy evaluations by taking a historical lens to try to understand these successful reformers. Future research could also usefully refine our measurement methods. Literacy is a crude indicator for overall learning levels, and it would be useful to corroborate the trends we measure with additional datasets containing more fine-grained learning metrics.

,

References

Helen Abadzi. *Improving Adult Literacy Outcomes Lessons from Cognitive Research for Developing Countries*, volume 0932. 2003. ISBN 0821354930.

Nadir Altinok and H Murselli. International database on human capital quality. *Economics Letters*, 96(2), 2007.

Nadir Altinok, Noam Angrist, and Harry Patrinos. Global data set on education quality (1965-2015). *World Bank Policy Research Working Paper 8314*, (January):82, 2018.

Tahir Andrabi, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics*, 3(3):29–54, 2011.

Noam Angrist, Simeon Djankov, Pinelopi K Goldberg, and Harry A Patrinos. Measuring human capital using global learning data. *Nature*, 592(7854):403–408, 2021.

Julius F Atuhurra. Does community involvement affect teacher effort? Assessing learning impacts of Free Primary Education in Kenya. *International Journal of Educational Development*, 49:234–246, 2016.

Robert Barro and Jong-Wha Lee. International Trade, Distortions and Long-Run Economic Growth. *IMF Working Papers*, 92(90):i, 2013.

Robert J Barro and Jong-Wha Lee. International comparisons of educational attainment. *Journal of Monetary Economics*, 32(3):363–394, 1993.

Sonia Bhalotra, Kenneth Harttgen, and Stephan Klasen. The impact of school fees on schooling outcomes and the intergenerational transmission of education. *UNESCO, Background paper for the Education for All Global Monitoring Report*, 2014.

Moussa P Blimpo, Ousman Gajigo, and Todd Pugatch. Financial constraints and girls' secondary education: Evidence from school fee elimination in The Gambia. *The World Bank Economic Review*, 33(1):185–208, 2019.

Tessa Bold, Mwangi Kimenyi, Germano Mwabu, and Justin Sandefur. Can free provision reduce demand for public services? evidence from kenyan education. *The World Bank Economic Review*, 29(2):293–326, 2015.

David Card. The Causal Effect of Education on Earnings. *Handbook of Labor Economics*, 3:1801–1863, 1999.

Francesco Caselli. Accounting for cross-country income differences. *Handbook of Economic Growth*, 1:679–741, 2005.

Lee Crawfurd and Aisha Ali. The case for free secondary education. In Justin Sandefur, editor, *Schooling for All*, chapter 3. Center for Global Development, 2022.

Janet Currie and Tom Vogl. Early-life health and adult circumstance in developing countries. *Annual Review of Economics*, 5(1):1–36, 2013.

Hai Anh H. Dang and Paul W. Glewwe. Well Begun, but Aiming Higher: A Review of Vietnam's Education Trends in the past 20 Years and Emerging Challenges. *Journal of Development Studies*, 54(7):1171–1195, 2018.

Jishnu Das and Tristan Zajonc. India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics*, 92(2):175–187, 2010.

Angus Deaton. *The analysis of household surveys: a microeconometric approach to development policy*. World Bank Publications, 1997.

Angus Deaton and Christina Paxson. Intertemporal Choice and Inequality. *Journal of Political Economy*, 102(3):437–467, 1994.

Klaus Deininger. Does cost of schooling affect enrollment by the poor? Universal primary education in Uganda. *Economics of Education Review*, 22(3):291–305, 2003.

Ethan Fosse and Christopher Winship. Analyzing age-period-cohort data: A review and critique. *Annual Review of Sociology*, 45:467–492, 2019.

Michel Garenne. Trends in nutritional status of adult women in sub-Saharan Africa. *DHS Comparative Reports*, (27):xv–pp, 2011.

Douglas Gollin, David Lagakos, and Michael E Waugh. The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993, 2014.

Kathleen S Gorman and Ernesto Pollitt. The contribution of schooling to literacy in Guatemala. *International Review of Education*, 43(4):283–298, 1997.

Jan Eric Gustafsson. Lasting effects of quality of schooling: Evidence from PISA and PIAAC. *Intelligence*, 57:66–72, 2016.

Giora Hanoch and Marjorie Honig. "True" age profiles of earnings: Adjusting for censoring and for period and cohort effects. *The Review of Economics and Statistics*, pages 383–394, 1985.

Eric A. Hanushek and Dennis D. Kimko. Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 90(5):1184–1208, 2000.

Eric A. Hanushek and Ludger Wossmann. Does education tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 2006.

S J Howie, C Combrinck, K Roux, M Tshele, G M Mokoena, and McLeod P N. *PIRLS Literacy 2016 : South African Highlights Report*. Centre for Evaluation and Assessment, Faculty of Education, University of Pretoria, 2017.

Aart Kraay. Methodology for a world bank human capital index. *World Bank Policy Research Working Paper*, (8593), 2018.

David Lagakos, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849, 2018.

N. G. Mankiw, D. Romer, and D. N. Weil. A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107(2):407–437, may 1992.

Karthik Muralidharan, Abhijeet Singh, and Alejandro J Ganimian. Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4):1426–60, 2019.

Una Okonkwo Osili and Bridget Terry Long. Does female schooling reduce fertility? evidence from nigeria. *Journal of development Economics*, 87(1):57–75, 2008.

Mari Oye, Lant Pritchett, and Justin Sandefur. Girls schooling is good, girls schooling with learning is better. *Background Paper for The Learning Generation. The Education Commission*, 2016.

Dev Patel and Justin Sandefur. A Rosetta Stone for Human Capital. *Center for Global Development Working Paper 550*, 2020.

Lant Pritchett. *The Rebirth of Education Why Schooling in Developing Countries is Flailing; How the Developed World is Complicit; and What to do Next.* Center for Global Development, 2013.

Stephanie R. Psaki, Erica Soler-Hampejsek, Jyotirmoy Saha, Barbara S. Mensch, and Sajeda Amin. The Effects of Adolescent Childbearing on Literacy and Numeracy in Bangladesh, Malawi, and Zambia. *Demography*, 56(5):1899–1929, oct 2019.

Sara J Ruto and John K Mugo. Free primary education in Kenya . The elusive EFA dream - about to be reached ? *Zeitschrift fur internationale Bildungsforschung und Entwicklungspadagogik 28 (2005) 1, S. 21-23*, 28, 2005.

Pasi Sahlberg. Education policies for raising student learning: The Finnish approach. *Journal of Education Policy*, 22(2):147–171, 2007.

Todd Schoellman. Education quality and development accounting. *Review of Economic Studies*, 79(1):388–417, 2012.

Abhijeet Singh. Learning More with Every Year: School Year Productivity and International Learning Divergence. *Journal of the European Economic Association*, 2019.

Erica Soler-Hampejsek, Barbara S. Mensch, Stephanie R. Psaki, Monica J. Grant, Christine A. Kelly, and Paul C. Hewett. Reading and numeracy skills after school leaving in southern Malawi: A longitudinal analysis. *International Journal of Educational Development*, 59:86–99, mar 2018.

Isabella Sulis, Francesca Giambona, and Mariano Porcu. Adjusted indicators of quality and equity for monitoring the education systems over time. Insights on EU15 countries from PISA surveys, 2020.

Stephen Taylor and Nicholas Spaull. The effects of rapidly expanding primary school access on effective learning: The case of Southern and Eastern Africa since 2000. *Stellenbosch Economic Working Papers: 1/13*, 2013.

Katarina Tomasevski. *The State of the Right to Education Worldwide. Fee or Free: 2006 Global Report*. 2006. URL http://www.katarinatomasevski.com/images/Global_Report.pdf.

UNESCO. The global learning crisis: Why every child deserves a quality education. Technical report, 2013.

Jan Van Damme and Kim Bellens. *Countries Strive Towards More Quality and Equity in Education: Do They Show Success or Failure? Evidence from TIMSS 2003 and 2011, for Grade 4*. 2017.

B. M. Ware, E. O., Kornu, D. D. K., & Adusei. Enhancing the entrepreneurial skills among Ghanaian students, the bedrock of job creation (entrepreneurship education). *Journal of Emerging Trends in Economics and Management Sciences*, 9(3):169–179, 2018.

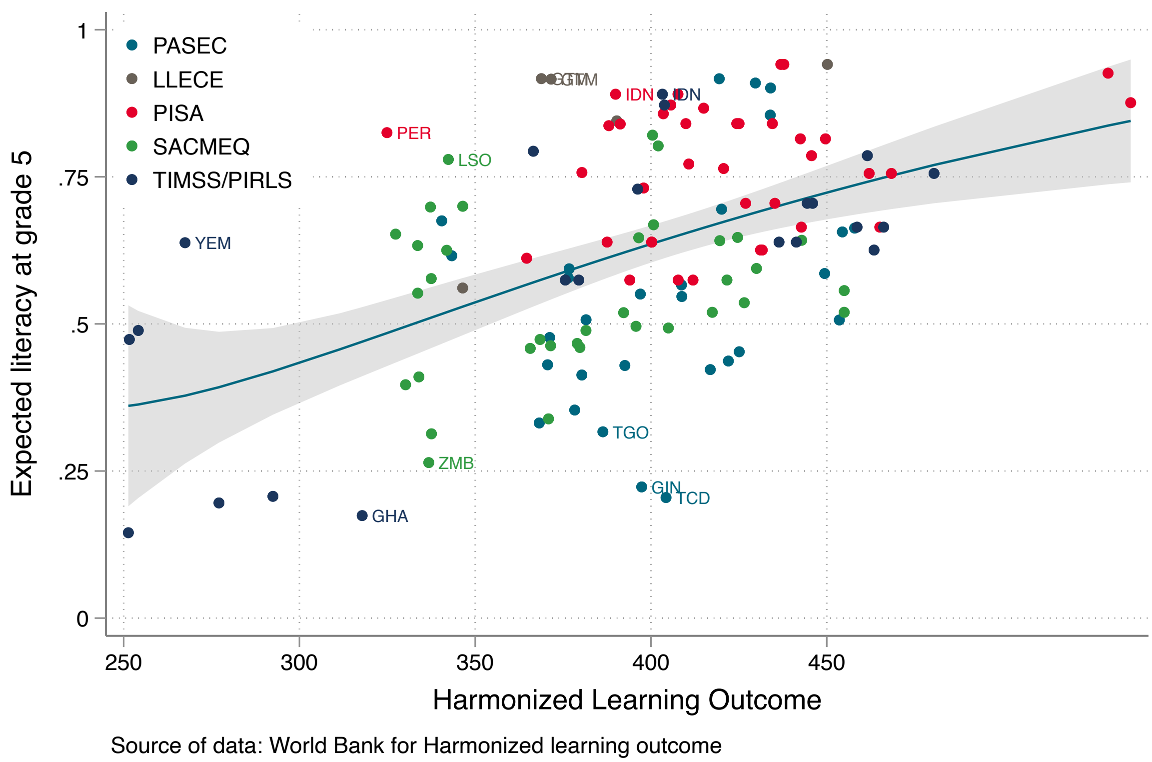
World Bank. *Learning: To Realize Education's Promise*. World Bank, 2018.

Ludger Wossmann and Eric Hanushek. The Role of Education Quality in Economic Growth The Role of School Improvement in Economic Development. *World Bank Policy Research Working Paper 4122*, pages 1–94, 2007.

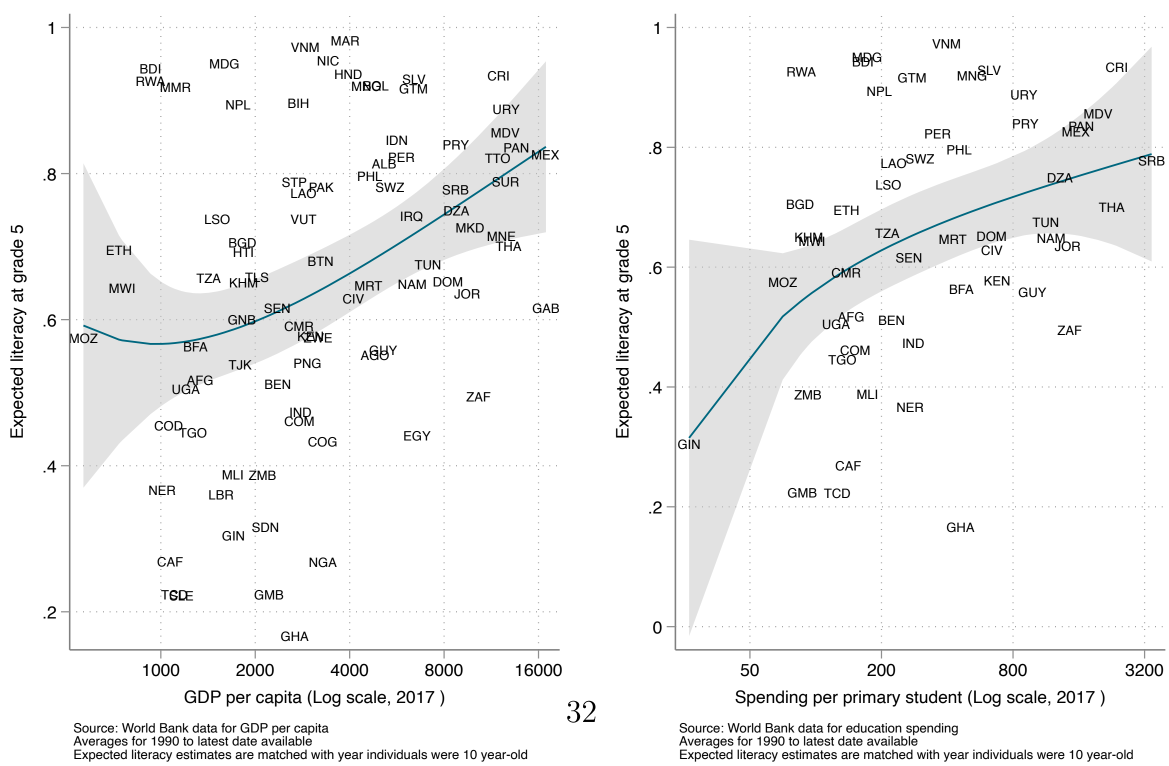
Yue Yin. Reflection on the Quality and Equity of Chinese Basic Education — From PISA 2018 Results : Where All Students Can Succeed. *Journal of Contemporary Educational Research*, 4, 2018.

Figure 1: Cross-sectional validation of education quality measure

(a) Education quality and harmonized learning outcomes

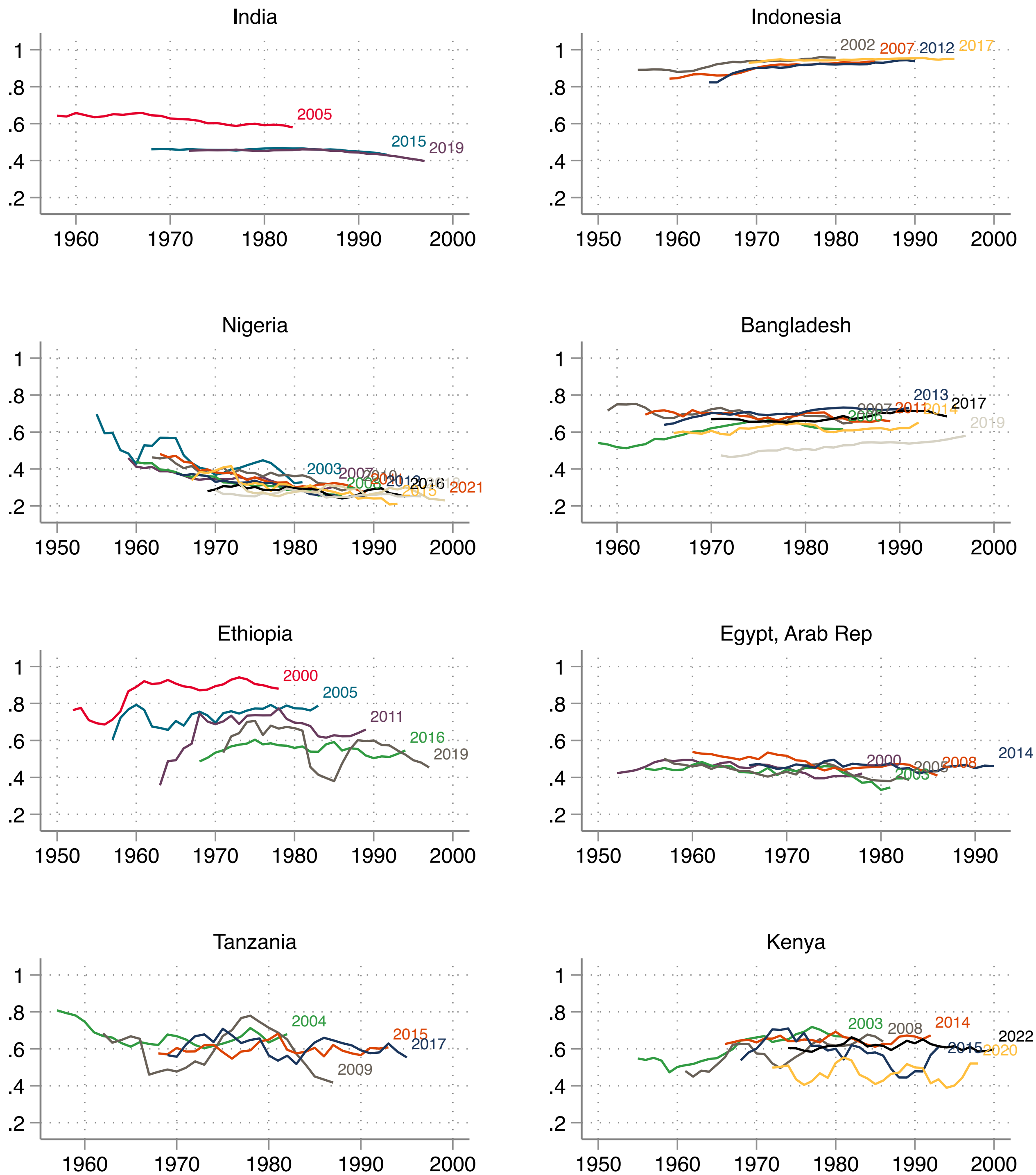


(b) Education quality, GDP per capita, and education spending



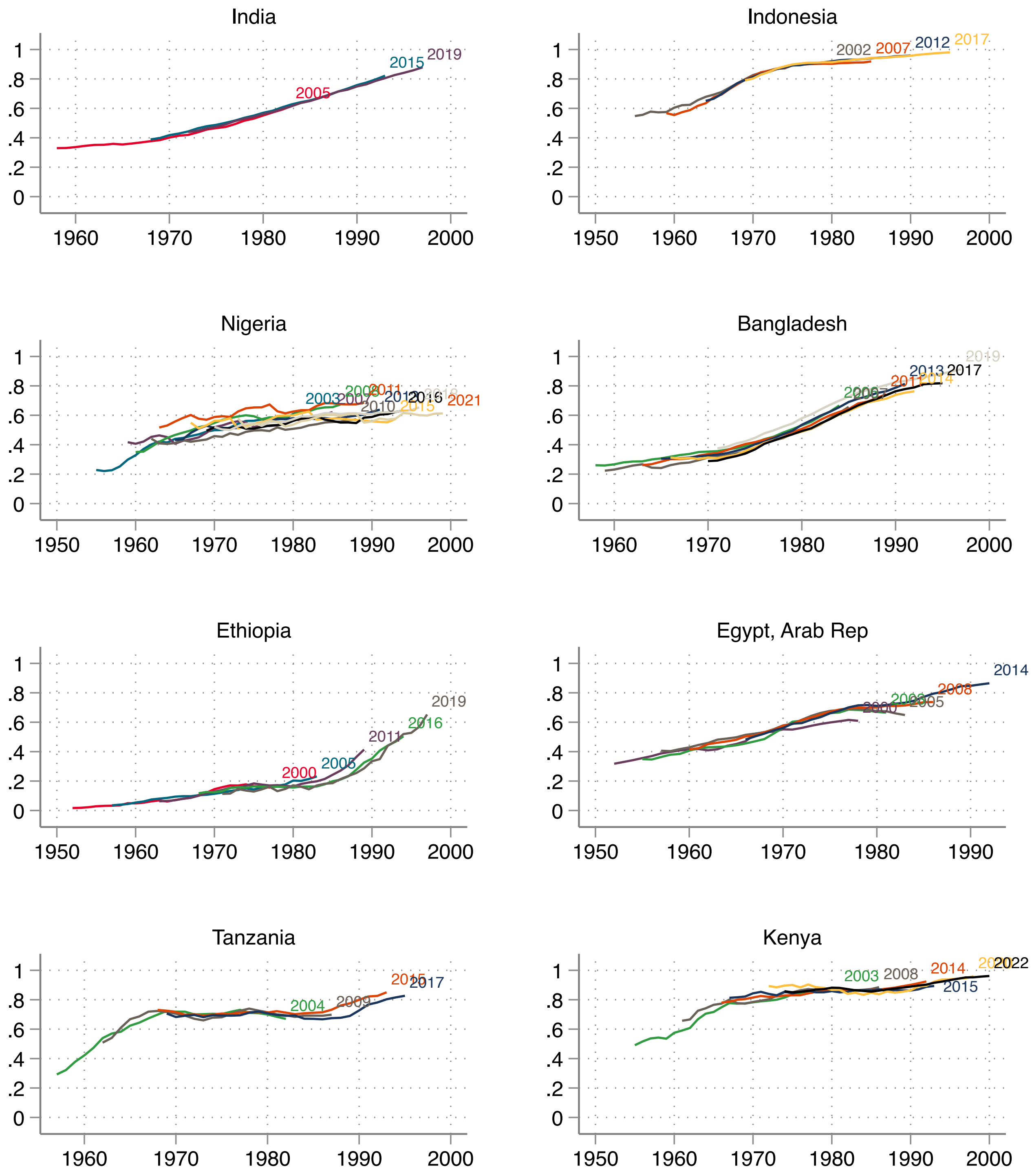
Note:

Figure 2: Literacy rates of women with four to six years of schooling (selected countries)



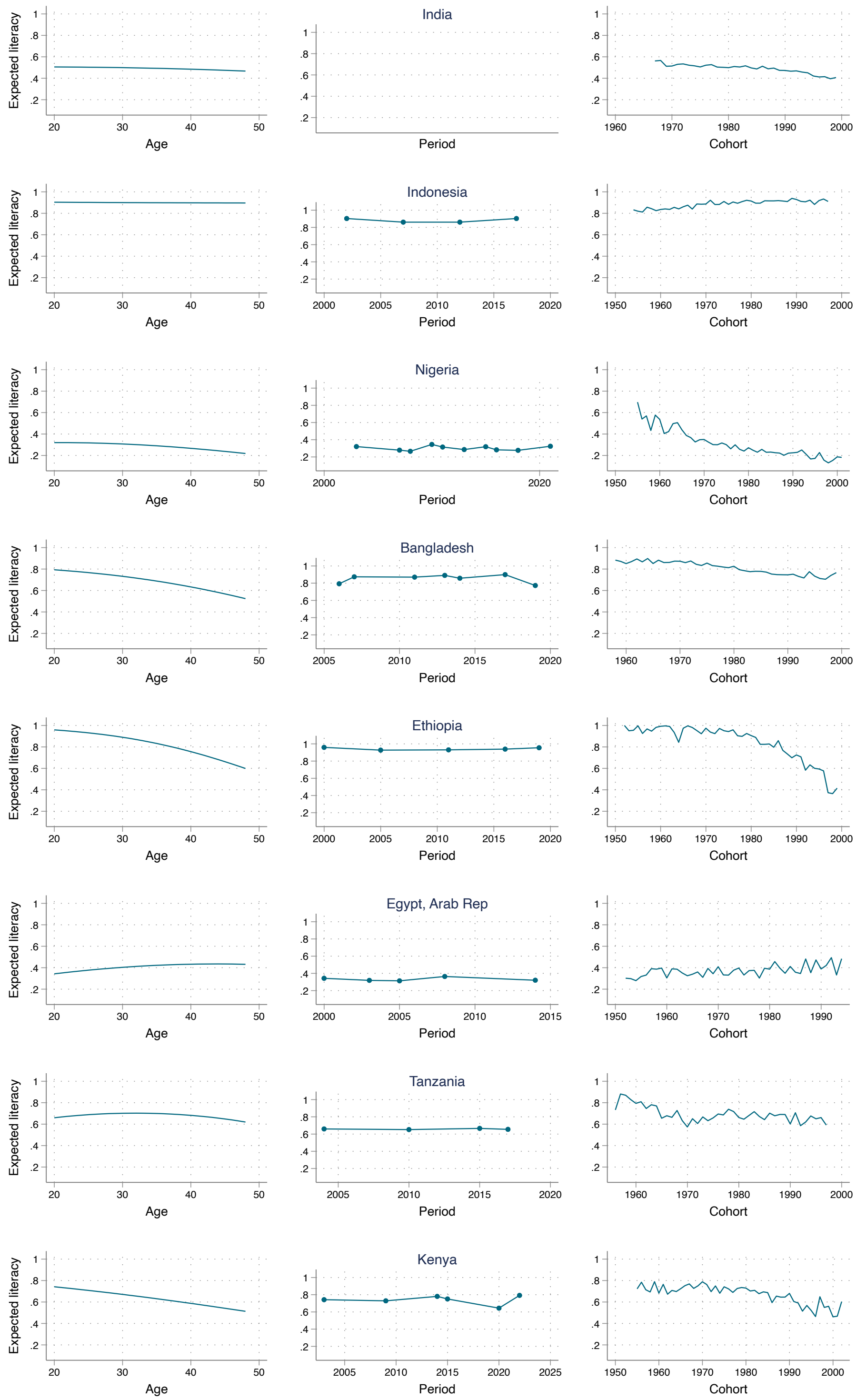
The vertical axis represents literacy rates 5-year moving average for literacy rates

Figure 3: Access to grade five for women (selected countries)



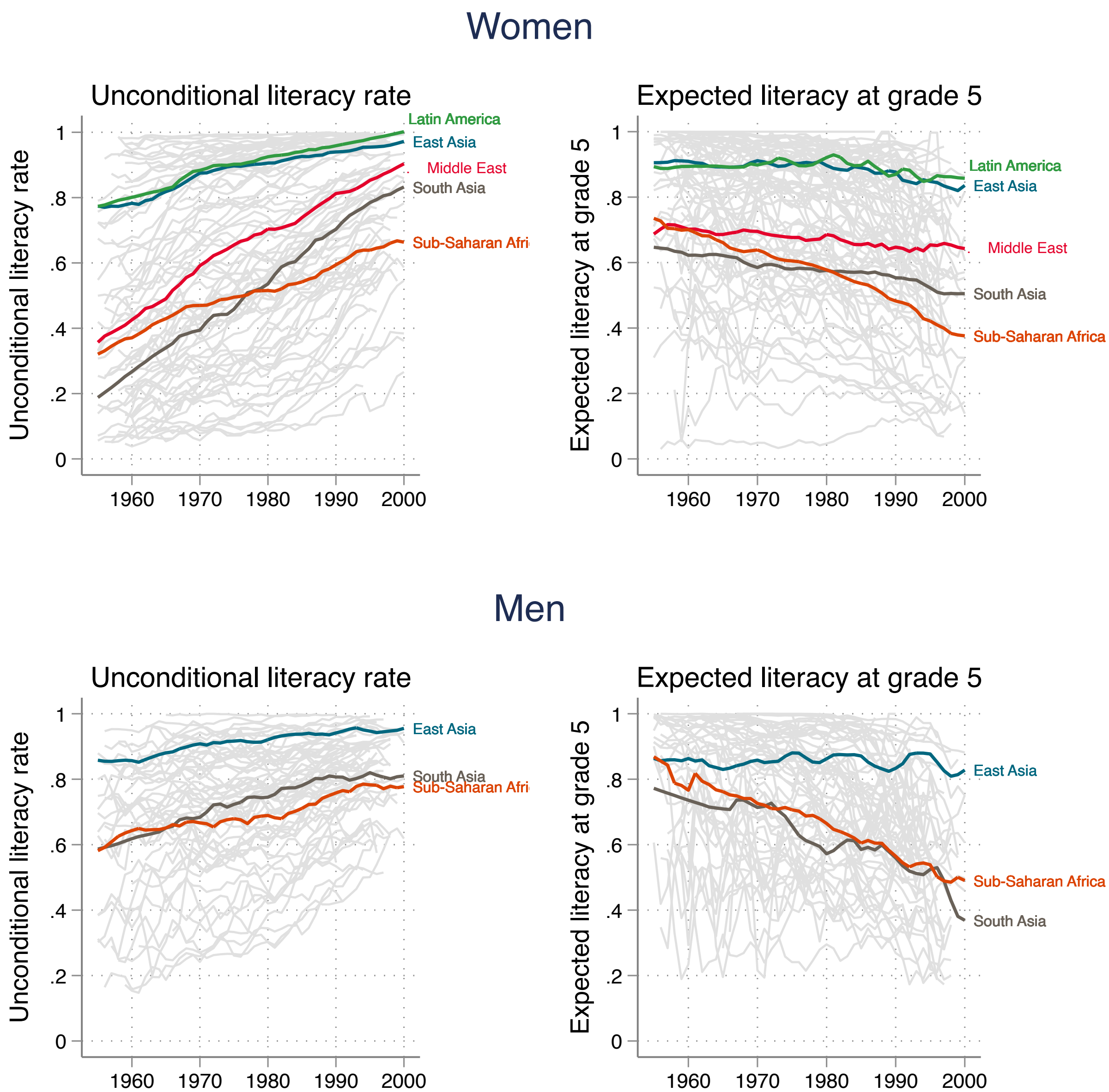
The vertical axis represents 5-year moving average of access to grade 5 or more

Figure 4: Estimated age, period and cohort effects in education quality (selected countries-women)



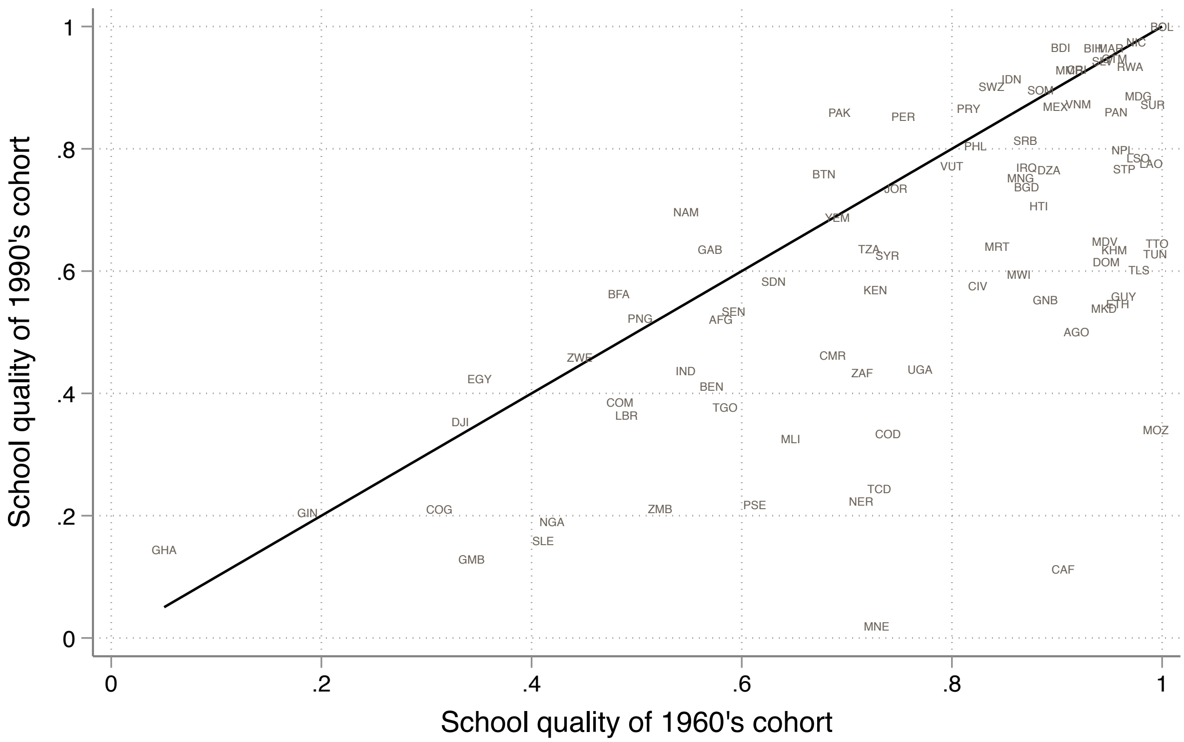
Outcome in all graphs is education quality, i.e. literacy conditional on five years of schooling. Age effects estimated at grade 5 for cohort born in 1975 and no period effect. Period effects estimated at age 20, grade 5 and 1975 cohort. Cohort effects estimated at age 20, grade 5 and no period effect.

Figure 5: National and regional trends in literacy and education quality



Note: Light gray lines show national estimates. To ensure regional trends reflect a consistent sample of countries, in cases of missing data for specific country-birth cohorts, national education quality estimates have been extrapolated backward and forward using the average, country-specific rate of change over the whole period.

Figure 6: Education quality for the 1960's cohort and 1990's cohort (women)



Note: 1980's cohort instead of 1990's for BIH, BOL, CAF, DJI, MAR, MKD, NIC, PHL, SOM and SYR.

Figure 7: Trends in height for all women and women with some schooling

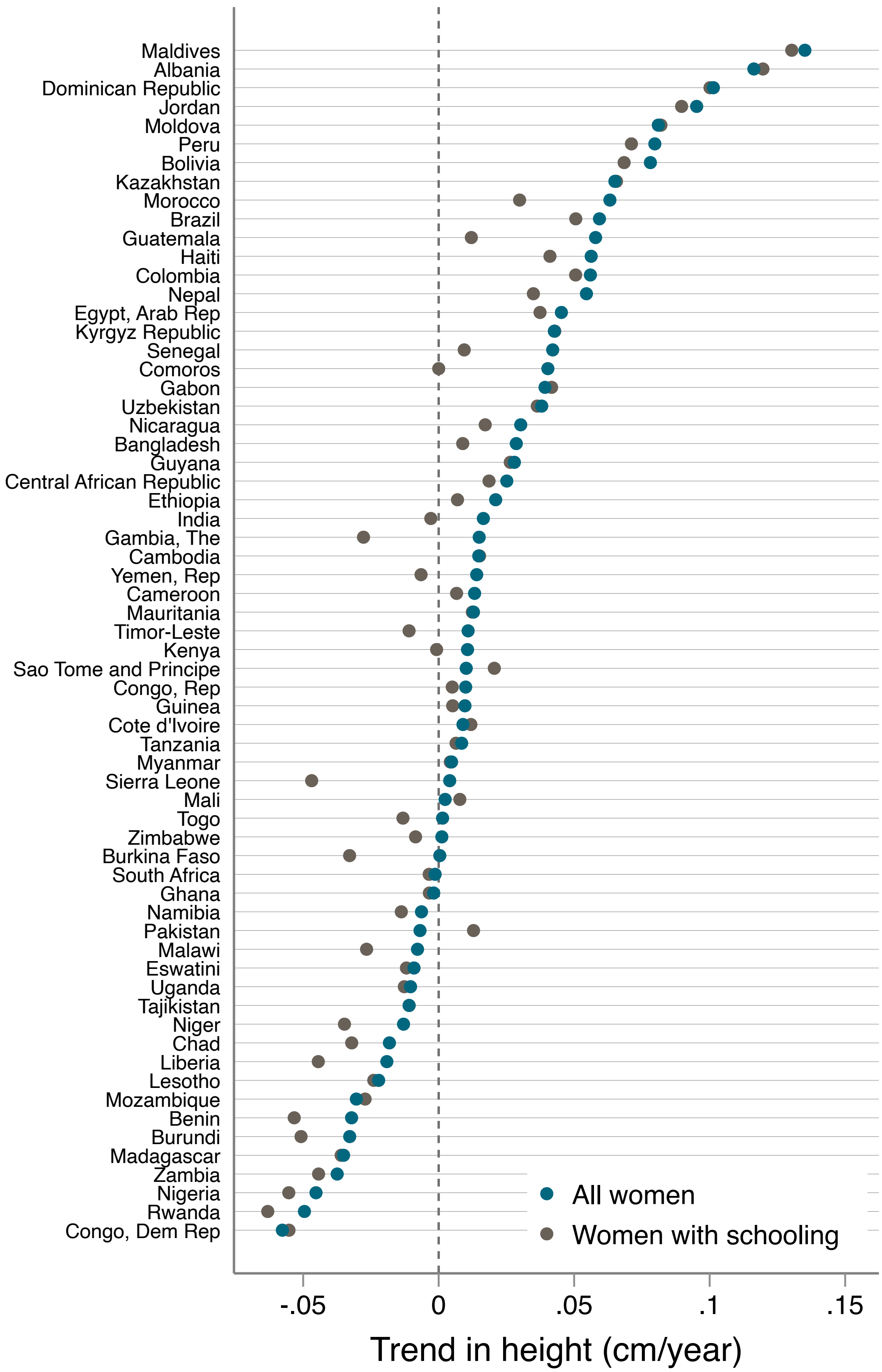
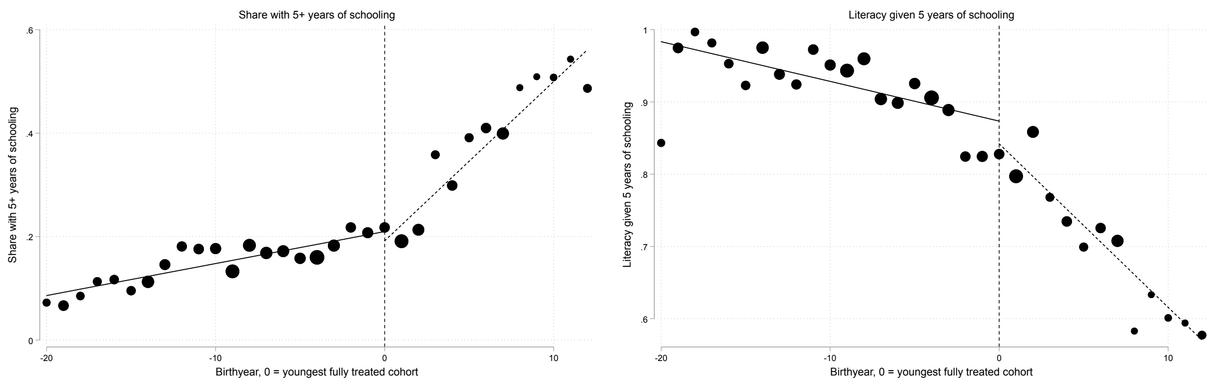
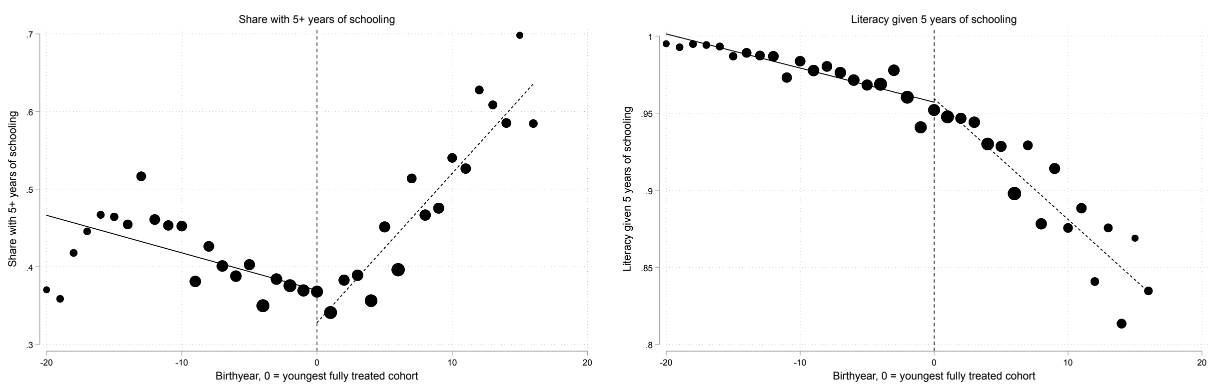


Figure 8: Trends in schooling and education quality before and after free primary education (FPE) reforms

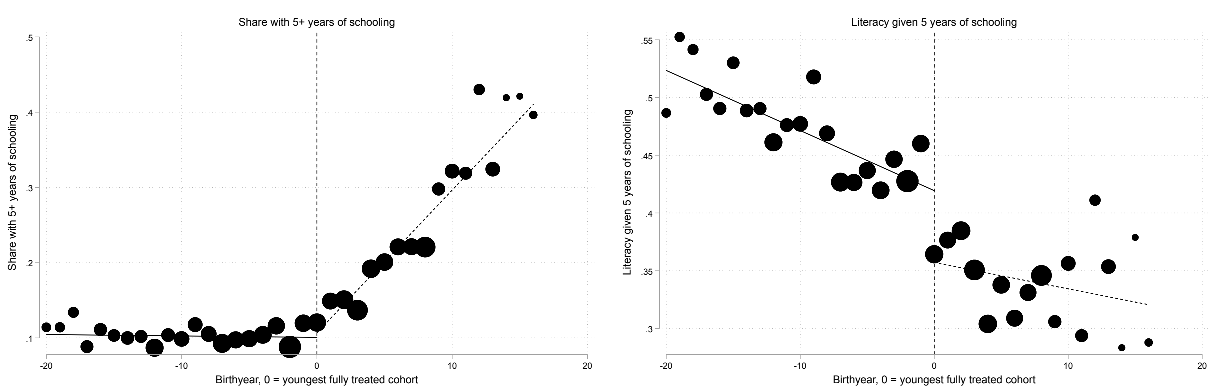
(a) Ethiopia, women



(b) Madagascar, women



(c) Mali, women



Note: The figure shows the cohort effects for schooling (share of women with five or more years of schooling) and education quality (literacy conditional on five years of schooling) from the model in equation (2). Vertical dashed lines mark the first cohort entering school after the abolition of fees. The three countries shown were selected on the basis of showing the largest trend breaks in women's schooling (rather than education quality).

Figure 9: Changes in education quality trends after free primary education (FPE) reforms

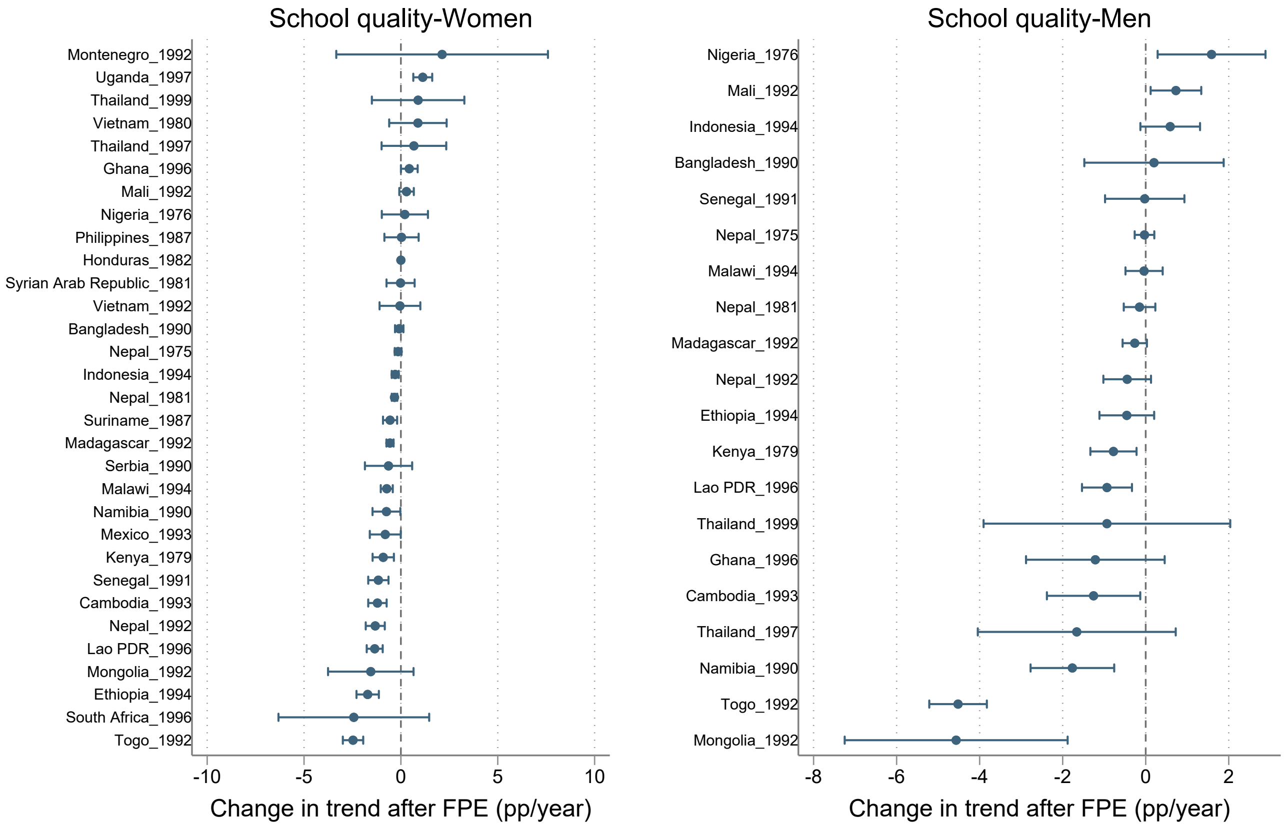


Table 1: Coverage

	Surveys	Countries	Countries with > 1 survey	Pop. share	Sample
Women					
East Asia and Pacific†	30	11	8	.91	529391
Europe and Central Asia	10	4	4	.03	47366
Latin America and Caribbean	33	14	8	.39	490365
Middle East and North Africa	18	10	4	.76	389545
South Asia	22	7	5	.99	1865404
Sub-Saharan Africa	189	41	38	.98	2210406
Total	302	87	67	.81	5532477
Men					
East Asia and Pacific†	17	10	4	.91	109592
Europe and Central Asia	3	2	1	.01	11767
Latin America and Caribbean	11	5	3	.08	82160
Middle East and North Africa	0	0	0	0	0
South Asia	13	6	4	.99	240623
Sub-Saharan Africa	125	37	36	.88	639881
Total	169	60	48	.69	1084023

Population share excludes high-income countries except Panama and Trinidad and Tobago.
† East Asia and Pacific excludes China.

Table 2: Literacy, share with at least five years of schooling and school quality

	Women				
	1950s	1960s	1970s	1980s	1990s
Literacy					
East Asia and Pacific	76.3	80.1	89.2	92.1	95.0
	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Latin America and Caribbean	71.3	85.8	90.2	94.0	96.5
	(0.10)	(0.04)	(0.03)	(0.02)	(0.01)
Middle East and North Africa	36.5	50.6	64.9	73.4	83.8
	(0.08)	(0.05)	(0.03)	(0.03)	(0.03)
South Asia	21.7	37.9	46.5	62.1	77.6
	(0.05)	(0.01)	(0.01)	(0.00)	(0.01)
Sub Saharan Africa	30.6	43.3	49.4	54.3	63.8
	(0.07)	(0.03)	(0.02)	(0.01)	(0.02)
Share with $S \geq 5$					
East Asia and Pacific	65.4	69.4	83.3	87.9	93.5
	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)
Latin America and Caribbean	57.0	77.0	82.4	89.4	94.9
	(0.12)	(0.05)	(0.04)	(0.03)	(0.02)
Middle East and North Africa	34.0	48.1	62.2	70.5	83.3
	(0.08)	(0.05)	(0.03)	(0.03)	(0.03)
South Asia	18.3	36.8	45.5	62.1	78.3
	(0.05)	(0.01)	(0.01)	(0.00)	(0.01)
Sub Saharan Africa	27.3	42.4	50.2	55.7	67.4
	(0.07)	(0.03)	(0.02)	(0.01)	(0.02)
School quality					
East Asia and Pacific	86.8	86.9	90.3	88.5	85.0
	(0.07)	(0.05)	(0.07)	(0.06)	(0.08)
Latin America and Caribbean	83.4	89.2	90.5	90.2	85.3
	(0.24)	(0.10)	(0.08)	(0.07)	(0.11)
Middle East and North Africa	62.2	69.7	67.8	66.4	59.4
	(0.17)	(0.11)	(0.09)	(0.09)	(0.16)
South Asia	76.8	59.6	58.5	57.0	52.7
	(0.37)	(0.04)	(0.03)	(0.02)	(0.02)
Sub Saharan Africa	70.2	66.3	60.7	53.7	42.6
	(0.17)	(0.08)	(0.06)	(0.06)	(0.05)
Countries					
East Asia and Pacific	7	10	11	11	10
Latin America and Caribbean	7	14	14	14	12
Middle East and North Africa	5	10	10	10	7
South Asia	3	7	7	7	7
Sub Saharan Africa	36	41	41	41	40

Table 3: Literacy, share with at least five years of schooling and school quality

	Men				
	1950s	1960s	1970s	1980s	1990s
Literacy					
East Asia and Pacific	86.5	87.3	91.3	93.3	94.0
	(0.05)	(0.04)	(0.03)	(0.02)	(0.03)
South Asia	46.8	68.0	72.4	78.0	82.6
	(0.17)	(0.02)	(0.01)	(0.01)	(0.02)
Sub Saharan Africa	61.1	65.6	67.4	71.0	77.6
	(0.12)	(0.05)	(0.03)	(0.03)	(0.03)
Share with $S \geq 5$					
East Asia and Pacific	74.1	76.9	85.3	89.2	90.9
	(0.07)	(0.05)	(0.04)	(0.03)	(0.04)
South Asia	33.5	64.1	69.5	76.3	82.0
	(0.17)	(0.02)	(0.01)	(0.02)	(0.03)
Sub Saharan Africa	53.5	60.3	64.5	69.4	77.6
	(0.13)	(0.06)	(0.03)	(0.03)	(0.03)
School quality					
East Asia and Pacific	90.2	88.4	87.1	86.2	86.5
	(0.10)	(0.08)	(0.07)	(0.09)	(0.12)
South Asia	99.7	72.9	66.0	59.3	51.4
	(0.02)	(0.08)	(0.08)	(0.06)	(0.06)
Sub Saharan Africa	79.5	76.9	70.5	62.0	52.2
	(0.16)	(0.12)	(0.08)	(0.07)	(0.07)
Countries					
East Asia and Pacific	2	8	10	10	9
South Asia	2	6	6	6	6
Sub Saharan Africa	26	37	37	37	37

Table 4: Gender gap in school quality

	1960s	1970s	1980s	1990s	
Sub-Saharan Africa					
Gender gap	-11.34	-11.06	-9.96	-8.61	1.48
	(0.07)	(0.05)	(0.04)	(0.04)	(0.12)
	37	37	37	37	0
East Asia & Pacific					
Gender gap	-2.62	3.15	2.25	-.62	-9.49
	(0.05)	(0.05)	(0.05)	(0.08)	(0.30)
	8	10	10	9	0
South Asia					
Gender gap	-13.17	-7.74	-2.68	1.08	-21.64
	(0.05)	(0.04)	(0.03)	(0.03)	(0.23)
	6	6	6	6	0

Gender gap=female school quality- male school quality. Results weighted by population of the birth cohort

Table 5: Education quality and economic growth

VARIABLES	(1) 1990	(2) 1990	(3) 2000	(4) 2000	(5) 2010	(6) 2010
School quality	0.0277***	0.0180**	0.0359***	0.0276**	0.0447***	0.0353**
	(0.00953)	(0.00856)	(0.0123)	(0.0108)	(0.0160)	(0.0141)
Years of schooling	0.00488***		0.00492***		0.00444	
	(0.00116)		(0.00134)		(0.00283)	
Share accessing grade 5		0.0304***		0.0253*		0.0173
		(0.0110)		(0.0131)		(0.0239)
Log of initial GDP pc (WB)	-0.0129***	-0.0106***	-0.0139***	-0.0101**	-0.0123*	-0.00720
	(0.00362)	(0.00346)	(0.00400)	(0.00406)	(0.00698)	(0.00618)
Constant	0.0599**	0.0647***	0.0628***	0.0617**	0.0490	0.0371
	(0.0225)	(0.0216)	(0.0226)	(0.0235)	(0.0359)	(0.0345)
Observations	61	74	66	83	67	85
R-squared	0.323	0.209	0.302	0.171	0.137	0.082

Note: The dependent variable is real economic growth per capita in constant USD from the initial year to 2019, as reported by the World Bank, World Development Indicators. Education in single years for adults is taken from Barro and Lee (2018). Education quality and share accessing grade five are the averages at initial year for the population aged 20 to 40 (or the highest available age) weighted by cohort population. Robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Education quality and enrollment

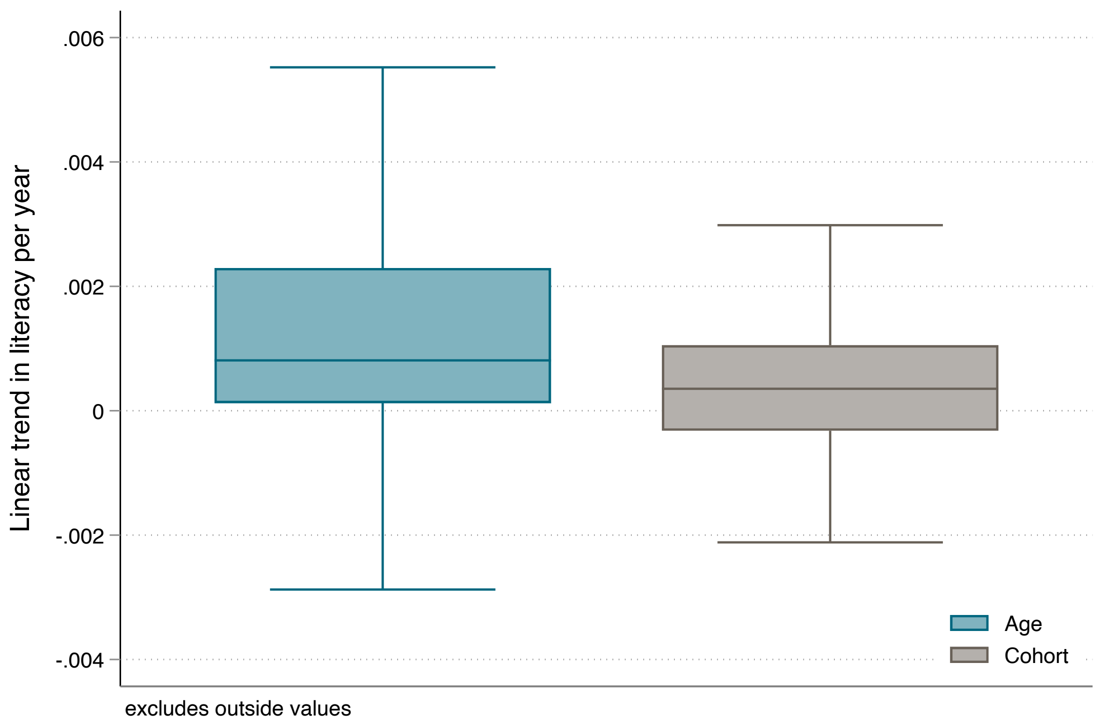
	(1)	(2)	(3)	(4)
Women				
Birth Year	-0.516***	-0.462***	-0.520***	-0.400***
	(0.0232)	(0.0530)	(0.0531)	(0.0467)
Share with schooling		-4.201	-1.972	8.482**
		(3.549)	(4.282)	(4.216)
Share with secondary schooling		-2.002	1.347	-11.97***
		(3.850)	(3.978)	(3.967)
Repetition			13.45***	8.305***
			(1.419)	(1.296)
Height w/ schooling				1.010***
				(0.338)
N	3326	3326	2275	1456
R^2	0.783	0.783	0.846	0.903
Men				
Birth Year	-0.595***	-0.576***	-0.714***	
	(0.0276)	(0.0474)	(0.0491)	
Share with schooling		15.74**	16.52**	
		(6.565)	(7.148)	
Share with secondary schooling		-11.37**		
		(4.635)		
Repetition			6.363***	
			(1.411)	
N	2288	2288	1638	
R^2	0.710	0.712	0.753	

Note: Each column presents a separate regression, estimated separately for men and women, of predicted literacy at grade 5 on the covariates shown in the left column. The unit of observation is a country-birth cohort. All estimates control for country fixed effects.
* for p < 0.05, ** for p < 0.01, and *** for p < 0.001.

A Appendix

A.1 APC decomposition of literacy for individuals with no schooling

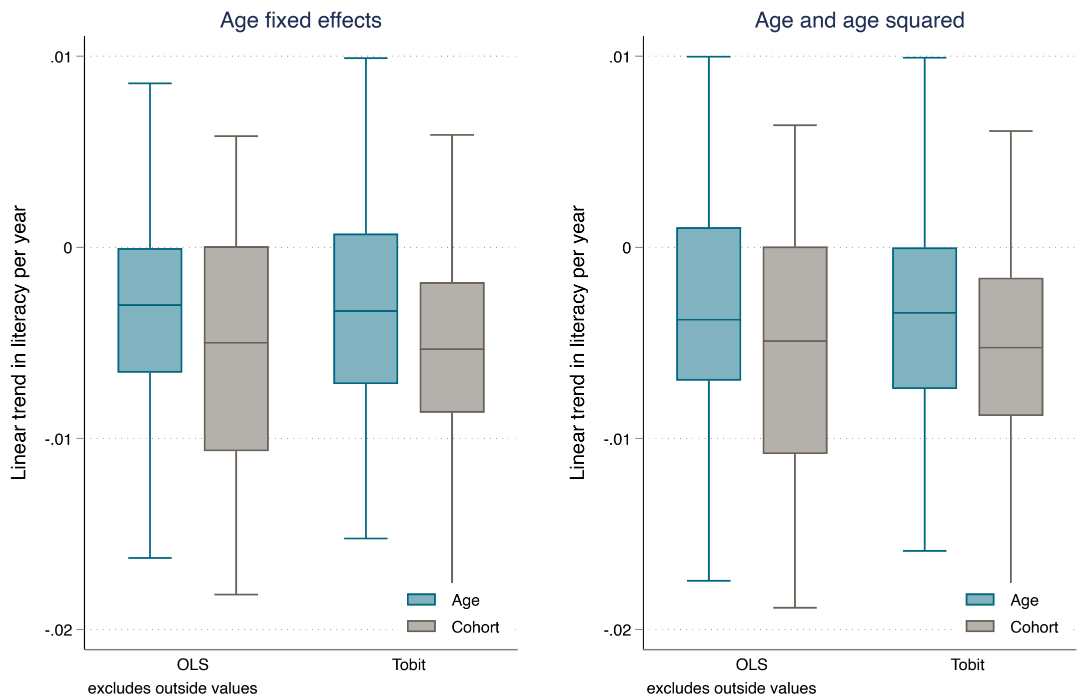
Figure 10: Age and cohort effects of literacy for individuals with no schooling. Deaton-Paxson normalization.



A.2 Age and cohort trends estimated with OLS and Tobit estimators

Results are summarized by looking at the linear trends of the age and cohorts as seen in and show OLS and Tobit estimates for equations 2 and 4. Period effects trends are not shown as they are zero by construction.

Figure 11: Comparison of age and cohort trends for OLS and Tobit estimates with age fixed effects and age and age squared



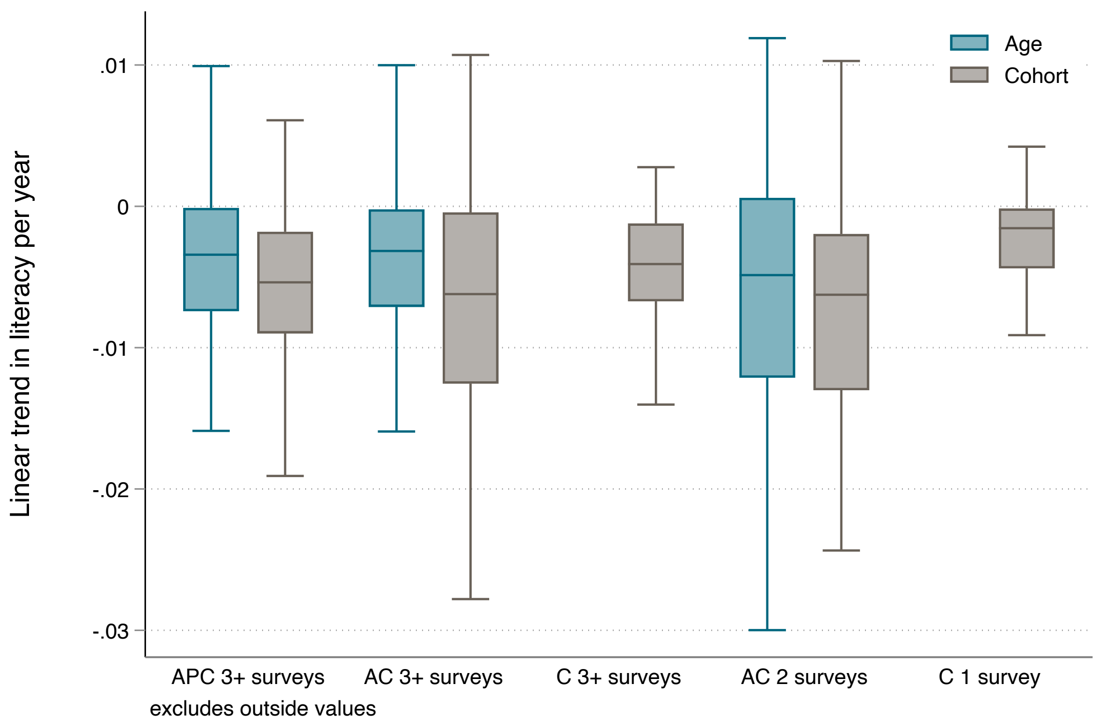
The different specifications show similar results. Age and cohort effects tend to be small and negative with about three quarters of estimated age and cohort trends being negative. Thus, both age and cohort effects tend to show that in the majority of cases individuals tend to lose literacy as they age and that more recent cohorts have lower rate of literacy after five years of schooling. In 12 cases, OLS and Tobit estimates using age fixed effects have estimation issues with parameters being omitted because of multicollinearity. Except for these 12 cases, age and cohort trends are almost identical when controlling for age fixed effects or age and age squared with a correlation of 0.99 for cohort effects and 0.9 for age effects for OLS and Tobit estimates. In order to keep a maximum number of cases, we therefore use the specification with age and age squared.

Compared to OLS estimates, Tobit estimates tend to have trends closer to 0 with a median age effect trend of -0.3 percentage points per year, or 9 percentage points over 30 years, while the median trend for age effect is -0.4 percentage points per year with OLS estimates. Similarly, median trends of cohort effects are -0.5 and -0.6 percentage points per year with

Tobit and OLS estimates respectively. The main difference lies in the trimming of large negative cohort effects. With Tobit estimates, there are 10 cases with cohort effects larger than -1 percentage point per year while there are 20 cases with OLS estimates. Interestingly, maximum values for positive cohort or age effects are similar with OLS or Tobit estimates. The trimming of large cohort effects makes results more plausible. For instance, for Ethiopian women OLS estimates predict a very large drop in school quality but this is due to the fact that expected literacy at grade five of 20 year old is 1.3 for the first cohorts and then decrease to about 0.7 for the last cohorts, whereas Tobit estimates predict a more realistic drop from 1 to 0.7. For this reason, we use Tobit estimates in our main discussion.

A.3 Comparison of APC, AC and C models

Figure 12: Comparison of APC, AC and C models by number of available surveys



For cases with at least three available surveys, dropping the normalized period effects (AC model) gives very similar results to the APC model estimated with DP normalization. Median trends in cohort and age effects are almost identical, which is not surprising given the

DP normalization assumes no trend in period effects. For comparison, age and cohort effects of AC estimates for country/sex cases with two surveys are shown in figure 12. They tend to show similar median trends than country/cases with more more than two surveys available but a wider distribution of effects. Finally, estimating a C model (with no controls for age or period) on country/sex cases with at least three available surveys shows that, as expected, the median cohort effects tend to be biased upwards (median cohort effects of -0.3 pp instead of -0.5 pp with the full APC decomposition), which is consistent with age effects being negative. Thus, for country/sex cases with only one available surveys we cannot rule out that the small negative median cohort trend (-0.1 pp) might be due to the absence of controls for age effects.

A.4 The case of India

In this section, we take a closer look at trends in India, which merit special attention for two reasons. Firstly, India alone makes up 33 percent of the underlying population represented by our 87-country sample of women. Secondly, the Indian literacy data show a strange pattern of changes across survey rounds. The basic conclusion of this appendix is that we should drop the earliest available round of Indian literacy data from the analysis. The result is that we estimate a gentler long-run decline in school quality for India, especially for women, than if we were to use all available data—though we do find a decline in all cases.

Our literacy data for India come from three rounds of the National Family Household Survey: NFHS 3 (2005/6), NFHS 4 (2015/6), and NFHS 5 (2019-21). Previous rounds did not include direct measurement of literacy. An earlier draft of this paper reported results with just the first two rounds, NFHS 3 and NFHS 4. [19] As we noted at the time, estimating trends using only two rounds of data requires strong assumptions—specifically, that there are no period effects, e.g., no changes in how literacy is measured across survey rounds. The public release of NFHS 5 allows us to re-examine this assumption. Three rounds are sufficient to apply the Deaton-Paxson method and generate estimates of age, period, and cohort effects, but remain somewhat sensitive to idiosyncracies of specific survey rounds compared to other countries with more rounds of data. Here we explore that sensitivity.

Our previous estiamtes using just two rounds of data (NFHS 3 and NFHS 4) showed a dramatic decline in school quality for India. Technically, those steep downward trends remain once adding the NFHS 5 data to the sample.

To understand how, in the most mechanical sense, the Indian results yield such steep downward trends it is useful to look at the raw data. Figure 2 in the main text shows the literacy rate of women and men with four to five years of primary of schooling by years of birth for NFHS 3 through 5 – without any allowance for age or period effects, or any other regression adjustment of any kind. For women, literacy rates are fairly flat aross birth cohorts within each survey, but there is a drop in literacy of more than 10 percentage points between NFHS 3 and NFHS 4 for women born the same year. (NFHS 5 shows no additional drop, with the literacy line falling almost exactly on top of the NFHS 4 line. The picture is similar for men although literacy rates are also declining within each survey.)

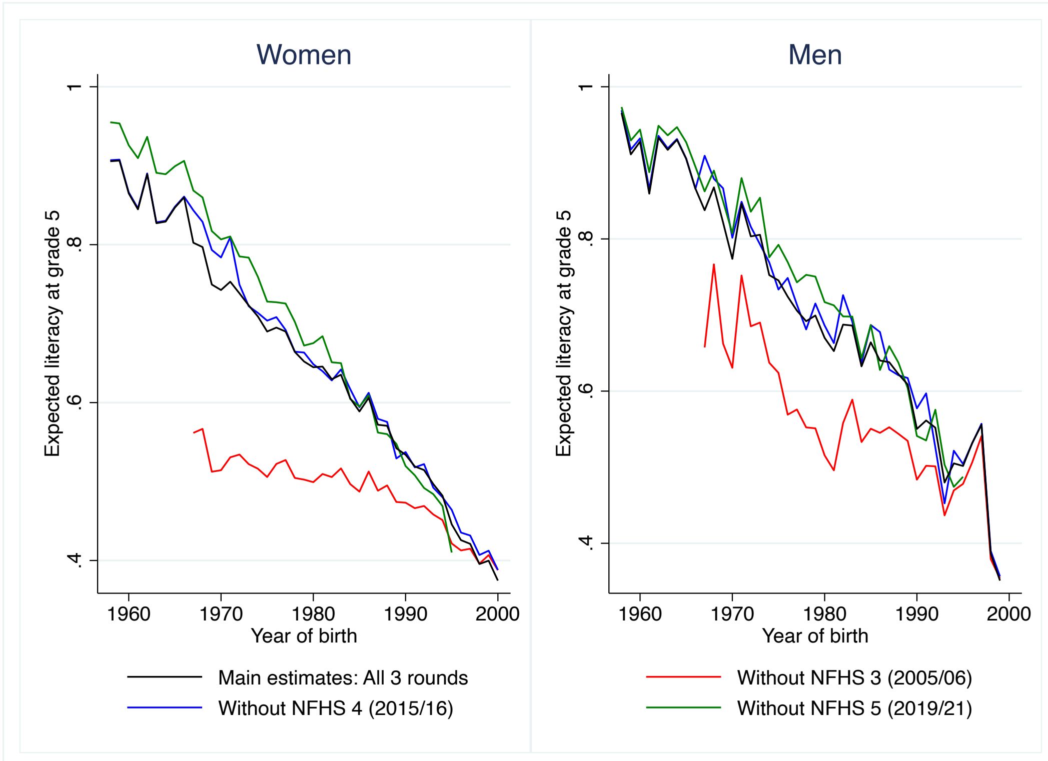
There are two ways to interpret the drop in literacy for women in the same birth cohort over the decade between NFHS 3 and NFHS 4: as age effects (literacy decayed over the life cycle) or as period effects (e.g., the surveys differed in terms of sampling or field administration). That choice has a direct implication for the implied cohort effects. If the discrepancy is due to age effects, then the flat literacy rates across cohorts in a given survey round actually reflect originally higher literacy for the older women, and we would conclude literacy at a given age is declining across cohorts. If, on the other hand, the discrepancy is due to period effects, then the flat lines we see in the raw data essentially tell us the trend across cohorts as well, and we would conclude literacy (conditional on schooling) has been quite flat.

The first interpretation is complicated by the fact that we see no further evidence of age decay in the four or five years between NFHS 4 and 5. It is hard to imagine why women across all age cohorts experienced a roughly 15 percent decline in literacy between 2005/6 and 2015/16, and then no additional decline by 2019-21. The second interpretation, that these discrepancies are due to survey round idiosyncracies, forces us to choose which surveys to believe and which rounds to discount. Retaining NFHS 3 together with either round 4 or 5 would imply a declining in (conditional) literacy across cohorts, while retaining only rounds 4 and 5 would imply a flat line.

To illustrate this, we reestimate age and cohort effects using only two rounds of data. We repeat this for all three possible pairs of NFHS rounds, as show in Figure 13. Dropping either NFHS 4 or 5 makes essentially no difference to the long-term trend in cohort effects: expected literacy at age 20 with 5 years of schooling is declining steadily over time for both men and women. In contrast, dropping NFHS 3 yields much flatter trends. The flatter trend stems from a much lower starting point: implied literacy rate (at age 20 with 5 years of schooling) for the 1970 birth cohort falls by roughly 20 percentage points for women and 10 percentage points for men.

[19] The February 2022 draft is archived here https://www.cgdev.org/sites/default/files/long-run-decline-education-quality-developing-world.pdf.

Figure 13: Robustness of India cohort effects to dropping individual survey rounds

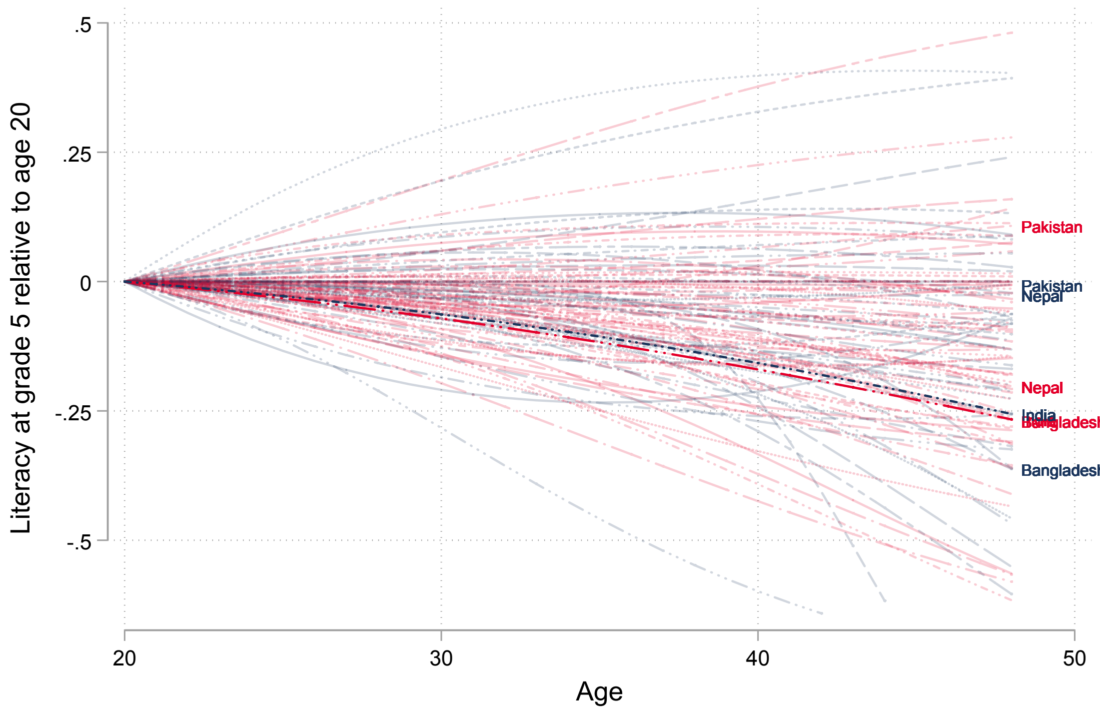


To repeat, the results from different samples in Figure 13 can be thought as two extreme modelling assumptions: either attributing all systematic differences between the various surveys to age effects or attributing them fully to period/survey effects. In both cases, education quality is declining but the magnitude of this decrease is quite different, especially for women and without additional data, it is not possible to know what modelling assumption is the most appropriate. However, there are reasons to think that our original modelling assumption controlling for age effects best fits the data.

Another possible source of survey effect could be a change in the way literacy is measured. In both surveys, individuals have to read a sentence and are graded on a 0-1-2 scale. The way interviewers grade literacy could have changed but results remain the same if people who are graded as 1 are classified as literate or illiterate. Also, differences in literacy rates are not due to differences in response rates that are high (>99 percent) and identical between the two surveys.

Estimated age effects for India are large but, as can be seen in figure 14, they are not outliers. They are similar to the estimated age effects for Bangladesh which have been estimated with seven surveys and also to those for Nepalese women estimated with six surveys. Thus, it is not impossible that there are indeed large age effects in India and that the more pessimistic estimates of a large drop in school quality, based on all three rounds of NFHS, are genuine.

Figure 14: Age effects for women (red) and men (blue)



Nevertheless, the estimated survey effect [SPECIFY ROUND] is 13 percentage points for women and 9.7 for men. That is a large effect, placing the discrepancy between NFHS rounds at about the 97th percentile in terms of the magnitude of survey effects in our sample. What could explain such a large discrepancy? First, we do not believe that there can be large 'true' period effects in adult literacy over a relatively short period of time. After allowing for normal life-cycle effects, it is unlikely that millions of Indian *adults* would have gained or lost literacy. over the space of a decade A sudden drop in participation in adult literacy programs is the type of explanation required here, but only 3.3 percent of women said they ever took part in a literacy program in the 2005/06 survey (the question was not asked in 2015/16).

The alternative explanation is survey effects, such as differences in sampling or data collection methods. All NFHS rounds are designed to be nationally representative and the sampling design was identical. However, the sampling frame changed between rounds 3 and 4 as there was a new census in 2011. Could this change explain the systematic difference between the two surveys? We think it is unlikely. First, similar differences between NFHS rounds 3 and 4 are found within urban and rural areas. Moreover, analysis by Indian state shows that a negative school quality trend and negative age effects are found in most states (see Figure 15), showing that the difference in literacy rates between the two surveys is not driven by composition effects. We also saw in Figure 3 in the main text that the average number of years of schooling for men and women is identical across the two surveys for individuals born the same year, suggesting there were not systematic differences in sampling between the two surveys.

Nevertheless, we have opted to report results using only the two most recent survey rounds in the main text. We have done this mostly to err on the side of caution—all options yield declining school quality in India, but restricting attention to NFHS 4 and NFHS 5 gives the least dramatic results. It is simply hard to look at literacy across the three rounds in Figure 2 and take the NFHS 3 results at face value. The implied survey effect from NFHS 3 is large, and we see no reason to explain age decay in literacy rates would have been so rapid in the 2000s and then stopped in the 2010s. The risk remains that we are discarding good data. The fact that the schooling measures match quite closely for each cohort across all NFHS rounds, and the implied age effects are well within the normal range internationally, both point in favor of using all three NFHS rounds, which would imply faster declines in school quality than we report in the main text. Hopefully future survey rounds will help resolve this uncertainty.

Figure 15: Trends in school quality by Indian state for women