WPS8591

POLICY RESEARCH WORKING PAPER 8591

BACKGROUND PAPER TO THE 2019 WORLD DEVELOPMENT REPORT

Learning-Adjusted Years of Schooling (LAYS)

Defining a New Macro Measure of Education

Deon Filmer
Halsey Rogers
Noam Angrist
Shwetlena Sabarwal

Abstract

The standard summary metric of education-based human capital used in macro analyses—the average number of years of schooling in a population—is based only on quantity. But ignoring schooling quality turns out to be a major omission. As recent research shows, students in different countries who have completed the same number of years of school often have vastly different learning outcomes. This paper therefore proposes a new summary measure, Learning-Adjusted Years of Schooling *(LAYS)*, that combines quantity and quality of schooling into a single easy-to-understand metric of progress. The cross-country comparisons produced by this measure are robust to different ways of adjusting for learning (for example, by using different international assessments or different summary learning indicators), and the assumptions and implications of *LAYS* are consistent with other evidence, including other approaches to quality adjustment. The paper argues that (1) *LAYS* improves on the standard metric, because it is a better predictor of important outcomes, and it improves incentives for policymakers; and (2) its virtues of simplicity and transparency make it a good candidate summary measure of education.

Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education*

Deon Filmer, Halsey Rogers,
Noam Angrist, and Shwetlena Sabarwal

This note proposes a new summary measure of education in a society: Learning-Adjusted Years of Schooling (*LAYS*). While simple in concept, this measure has the desirable property that it combines the standard macro metric of education—which captures only the quantity of schooling for the average person—with a measure of quality, defined here as learning. This adjustment is important for many purposes, because recent research shows that students who have completed the same number of years of school often have vastly different learning outcomes across different countries. While this adjustment may be meaningful even for comparisons of education in different high-income countries, it is especially important when we bring low- and middle-income countries into the comparative analysis, because the measured learning gaps between students become much larger.

The paper is structured as follows: Section 1 explains why we would want to adjust schooling for learning; Section 2 defines the *LAYS* measure; Section 3 discusses how to interpret *LAYS*; Section 4 explores the *LAYS* measure's robustness to different sources of learning data; Section 5 presents supporting evidence for the validity of the *LAYS* approach; Section 6 discusses using *LAYS* as a policy measure and briefly describes alternative approaches to adjusting years of schooling.

1. Why adjust schooling for learning?

Reliable macro measures of the amount of education in a society are valuable. First, they serve as metrics of progress: they allow a system to measure how well it is educating its people, and thus gauge the performance of education systems. Second, they are inputs for research and analysis: many empirical analyses of education's effects use aggregate schooling measures to explain variations in economic growth, productivity, health, governance quality, and other outcomes.

The typical proxy for education used in aggregate-level contexts is a quantity-based measure: the number of years of schooling that have been completed by the average member of the population (or sometimes by the average worker). This schooling-based measure does indeed predict some outcomes of interest—such as income and health—which is one reason it is widely used. But for reasons discussed below, an education measure that combines both quantity and quality of schooling may be preferable for many research and policy purposes.

2

1.1 Schooling is not the same as learning

Schooling is an imprecise proxy for education, because a given number of years in school leads to much more learning in some settings than in others. Or, to state it more succinctly, schooling is not the same as learning (Pritchett 2013, World Bank 2018). Recent studies make this very clear:

- International large-scale student assessments such as the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), and Progress in International Reading Literacy Study (PIRLS) reveal stark differences across countries in the levels of cognitive skills of adolescent students at the same age (for example, age 15 for PISA and 8th grade for TIMSS). In some participating countries, children's learning on average lags several years behind that of their peers in other countries.

- Other evidence more focused on middle- and low-income countries also shows wide gaps in learning across countries. In Nigeria, for example, 19 percent of young adults who have completed only primary education are able to read; by contrast, 80 percent of Tanzanians in the same category are literate. At any completed level of education, adults in some countries have learned much more than adults in other countries. (See Figure 1.1.)

Figure 1.1:   Literacy rates at successive education levels, selected countries
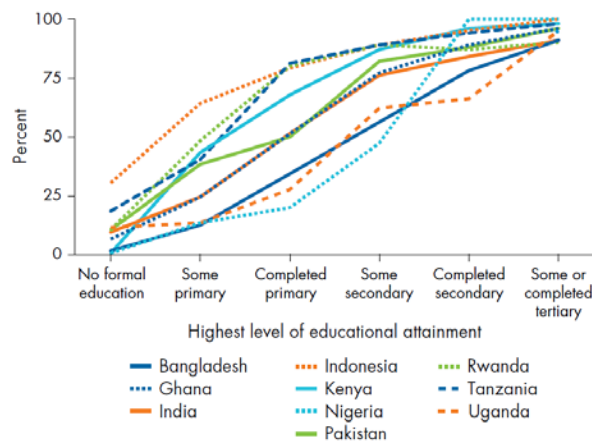
Source: Kaffenberger and Pritchett (2017), as reproduced in World Bank (2018).
Note: Literacy is defined as being able to read a three-sentence passage either "fluently without help" or "well but with a little help."

1.2 Learning matters

These learning gaps matter, because learning and skills drive many development outcomes. As the World Development Report (WDR) 2018 argues,

> Intuitively, many of education's benefits depend on the skills that students develop in school. As workers, people need a range of skills—cognitive, socioemotional, technical—to be productive and innovative. As parents, they need literacy to read to their children or to interpret medication labels, and they need numeracy to budget for their futures. As citizens, people need literacy and numeracy, as well as higher-order reasoning abilities, to evaluate politicians' promises. As community members, they need the sense of agency that comes from developing mastery. None of these capabilities flows automatically from simply attending school; all depend on learning while in school. (World Bank 2018, pp. 45-46)

Although the empirical literature on impacts of education has focused much more on schooling than on learning, mounting evidence supports this intuition. Even after controlling for schooling, empirical studies find that levels of learning and skills in the adult population affect outcomes:

- *Earnings of individuals:* "Across 23 OECD countries, as well as in a number of other countries, simple measures of foundational skills such as numeracy and reading proficiency explain hourly earnings over and above the effect of years of schooling completed" (WDR 2018, citing Hanushek and others 2015 and Valerio and others 2016).

- *Health:* Across 48 developing countries, "[e]ach additional year of female primary schooling is associated with roughly six fewer deaths per 1,000 live births, but the effect is about two-thirds larger in the countries where schooling delivers the most learning (compared with the least)" (WDR 2018, citing Oye, Pritchett, and Sandefur 2016).

- *Financial behavior:* **"**Across 10 low- and middle-income countries, schooling improved measures of financial behavior only when it was associated with increased reading ability" (WDR 2018, citing Kaffenberger and Pritchett 2017).

- *Social mobility:* In the United States, "the test scores of the community in which a child lives (adjusted for the income of that community) are among the strongest predictors of social mobility later in life**" (**WDR 2018, citing Chetty and others 2014), indicating that education quality has an impact beyond the number of school years completed.

- *Economic growth:* "[L]earning mediates the relationship from schooling to economic growth. While the relationship between test scores and growth is strong even after controlling for the years of schooling completed, years of schooling do not predict growth once test scores are taken into account, or they become only marginally significant" (WDR 2018, citing Hanushek and Woessmann 2012; see Figure 1.2).

The actual effects of learning may be even larger, for at least two reasons. First, the measures of learning used in the literature are necessarily incomplete, and sometimes very rough. For example, to obtain estimates of the learning effects on health across so many low- and middle-income-countries, the Oye, Pritchett, and Sandefur (2016) study cited above has to rely on just one very simple measure of skills: whether the respondent could read and understand a simple sentence such as "Farming is hard work." More sophisticated measures would likely explain more of the variation in outcomes. Second, learning has indirect effects that aren't captured in these estimates. The studies cited above all control for the number of years of schooling, but students with better cognitive skills are likely to stay in school longer, and at least some of this effect is likely causal. In some cases, a student who learns more will be able to persist longer in school for mechanical reasons, for example if it enables her to pass an examination to enter the next level of schooling. In other cases, learning more may keep the student from becoming frustrated with school and dropping out.

Figure 1.2: Correlations between two different education measures (test scores and years of schooling) and economic growth
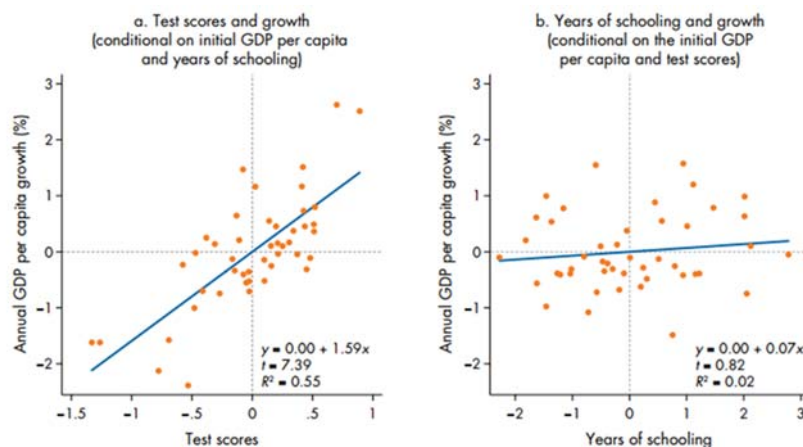
Source: WDR 2018, based on Hanushek and Woessmann (2012), using data on test scores from that study and data on years of schooling and GDP from World Bank's World Development Indicators

Beyond these instrumental benefits, improving learning matters if governments care about living up to the commitments they have made to their populations. Education ministries everywhere set standards for what children and youth are supposed to have learned by a given age, but students' learning often falls well short of what these standards dictate. For example, in rural India in 2016, a study found that only half of grade 5 students could fluently read text at the level of the grade 2 curriculum (ASER Centre 2017).

1.3 Adjusting the standard measure to reflect learning: the LAYS approach

Because it does not account for these differences in the learning productivity of schooling, the standard years-of-schooling approach to measuring education may be misleading, from both a policy and research perspective. In the policy world, for example, when the Millennium Development Goals' headline education measure targeted only the quantity of schooling (specifically, pledging to achieve universal primary completion by 2015), it created unintended incentives to discount schooling quality and student learning. From a research perspective, as the examples above show, measures that fail to incorporate quality will lead to underestimating education's benefits. The question, then, is how best to incorporate quality and learning outcomes into the standard macro measures, and thus enable more meaningful comparisons.

The approach described here is to adjust the standard years-of-schooling measure using a measure of learning productivity—how much students learn for each year they are in school. The WDR 2018 proposed such an adjustment and provided a simple illustration (World Bank 2018, Box 1.3). This note further develops that Learning-Adjusted Years of Schooling approach. As noted above, *LAYS* has the intuitively attractive feature that it reflects the quantity *and* quality of schooling, both of which societies typically view as desirable.[1] And by combining the two, it avoids the weaknesses of using either measure alone: unlike the years-of-schooling measure

---

[1] One might question why we should pay any attention to quantity-based schooling measures at all. In theory, we could simply use a measure of the learning and skills that a student leaves school with, and give no credit for the number of years spent in school. A rebuttal is that all skills measures are incomplete, and that schooling has other unmeasurable benefits that matter (and that are correlated with years of schooling).

alone, it keeps focus on quality; and unlike the test-score measure alone, it encourages schooling participation of all children, whether or not they will score highly on tests. The next section describes how *LAYS* is calculated.

2. Defining the LAYS measure

The objective of this exercise is to compare years of schooling across countries, while adjusting those years by the amount of learning that takes place during them. Ultimately, the measure we derive is defined as a quantity akin to:

$$LAYS_c = S_c \times R_c^n \tag{1}$$

where S_c is a measure of the average years of schooling acquired by a relevant cohort of the population of country c, and R_c^n is a measure of learning for a relevant cohort of students in country c, relative to a numeraire (or benchmark) country n. One straightforward way to define R_c^n is to use the highest-scoring country in a given year as the numeraire (meaning that R_c^n will be less than 1, for all countries other than the top performer), although as discussed below, we could establish this numeraire in other ways. For now, we define the measure of relative learning as:

$$R_c^n = \frac{L_c}{L_n} \tag{2}$$

where L_c and L_n are the measures of average learning-per-year in countries c and n respectively.[2] L can be thought of as a measure of the learning "productivity" of schooling in each country, and R is productivity in country c relative to that in country n. (As with the choice of numeraire, below we explore other possible ways of measuring relative learning.)

In the simplest sense, *LAYS* can be straightforwardly interpreted as an **index** equal to the product of two elements, **average years of schooling** and a **particular measure of learning relative to a numeraire**. Interpreting *LAYS* in this way requires no further assumptions or

---

[2] While education systems are clearly designed to produce outputs other than learning and test scores, this learning-adjustment exercise focuses on narrowly defined and measured outcomes.

qualifiers: it stands on its own and is clearly defined.[3]

The WDR 2018 illustrated this approach using: (1) the Grade 8 TIMSS learning assessment results for mathematics in 2015 to derive L_c; (2) mean years of schooling completed by the cohort of 25- to 29-year-olds, as calculated by Barro and Lee (2013) to measure years of schooling S_c; and (3) the learning achievement of Grade 8 students in Singapore (the top performer on this assessment) to derive L_n.[4] The resulting chart, which appeared in the WDR 2018, is reproduced here as Figure 2.1. Based on this calculation, for example, 25- to 29-year-olds in Chile have on average 11.7 years of schooling; the learning adjustment reduces that to 8.1 "adjusted" years. The same cohort in Jordan has 11.1 years of schooling on average; adjusting for learning brings that down to "adjusted" 6.9 years.

Figure 2.1: Average years of schooling of the cohort of 25- to 29-year-olds, unadjusted and adjusted for learning (using the *LAYS* adjustment)
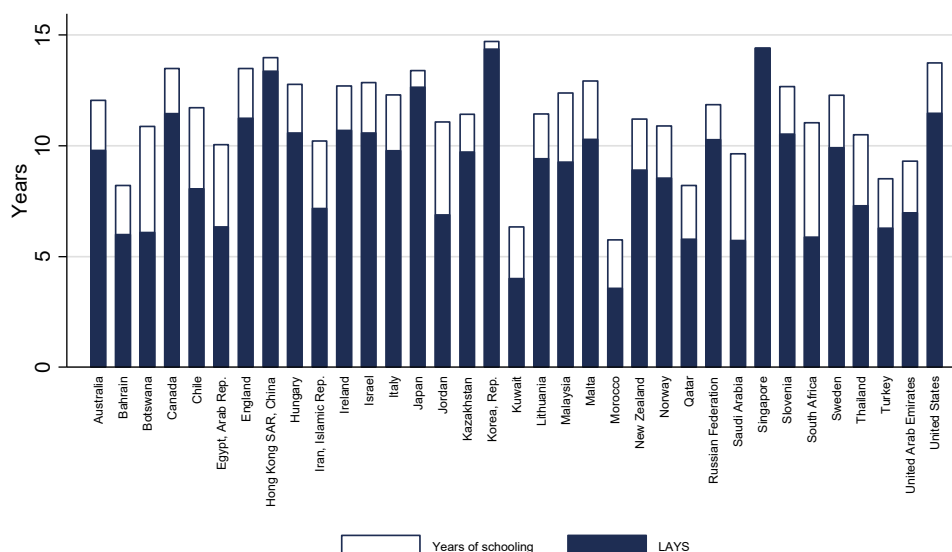
Source: WDR 2018 (World Bank 2018), based on analysis of TIMSS 2015 and Barro and Lee (2013) data.

---

[3] Of course, as a "mash-up" index, many other possible approaches to scaling and combing average years of schooling and learning outcomes are possible, e.g. using relative years of schooling, or adding rather than multiplying the two indicators. As will become clear in the next section, we do aim to provide a more substantive meaning to the index.

[4] The illustration included the additional assumption that learning starts at Grade 0, a point we come back to below.

3. Interpreting the *LAYS* measure

In this section, we first discuss how to interpret *LAYS*, explaining the mechanics and assumptions that underlie the learning adjustment. We then explore how different assumptions about when a child's learning starts—upon school enrolment, at birth, or somewhere in between—will affect the *LAYS* calculations.

3.1 Average learning profile

Assigning more meaning to *LAYS*—specifically, treating it as a **measure of years of schooling adjusted for quality**—requires interpreting the measure of learning and making certain assumptions. This is because internationally comparable measures of learning are typically tests (or assessments) administered at just one grade (or at one age, or one point in the schooling cycle, such as "end of primary" or "end of lower secondary"), whereas the *LAYS* measure typically applies the learning adjustment at another grade (that is, after a different number of years of schooling).

At any given grade, assessment scores measure students' **cumulative** learning up to that point. Therefore, the average annual "productivity" of an education system at producing learning up until that point is this learning measure divided by the number of years of schooling prior to the assessment.[5] In this case, L_c is therefore defined as

$$L_c = \frac{T_c}{Y_c} \tag{3}$$

where T_c is the test score and Y_c is the number of years of schooling preceding the assessment (i.e., the grade in which the assessment is administered).

As noted above, and we elaborate on this point below, the average years of schooling (which is the measure of schooling that we adjust) will generally differ from the number of years of schooling preceding the assessment (which is the measure of learning we're using to adjust schooling). In Figure 2.1, for example, we adjust for Chile's 12 years of schooling with a test

---

[5] In the next section, we discuss the implications of a potential difference between "years of schooling" and "years of learning"—that is, different assumptions about when learning begins.

that was administered in Grade 8. Therefore, we are using the average productivity measure across all grades, even though it may not apply directly to all those grades: the years it covers may not include the one we're adjusting (if the average number of years of schooling is greater than the years of schooling preceding the assessment), or it may reflect more years than the average number we're adjusting (if the average is less than the number of years of schooling preceding the assessment).[6]

This approach to calculating learning-adjusted years of schooling is illustrated graphically in Figure 3.1, using a hypothetical example. Assume that we observe Grade 8 test scores of 600 for Country A and 400 for Country B (illustrated as the red and blue points A and B) and that the average number of years of schooling in Country B of 9 years (illustrated as the vertical black line). The goal of the *LAYS* exercise is to "convert" the 9 years of schooling in Country B into the number of years of schooling in Country A that would have produced the same level of learning.

The conversion relies on the **average learning profiles** in the two countries, represented by the slopes of the lines from the origin to the points representing the observed test scores. Moving along the average learning profile from Grade 8 (for which we have the test score) allows us to infer what Country B's average score would be in Grade 9 (which, recall, is the average years of schooling for Country B in this example). This is represented by the move from point B to point C, or from a test score of 400 to 450. The next step is to go from point C to point D, to find the number of years of schooling that it would take in Country A to produce that level of learning (450), given the average learning profile in Country A. In this example, it takes just 6 years, so the resulting learning-adjusted years of schooling measure in Country B is 6.

---

[6] Assuming that learning rates are roughly constant across grades, this assumption will not be problematic. Below, we show that this is often the case.

Figure 3.1: Graphical illustration of deriving Learning-Adjusted Years of Schooling
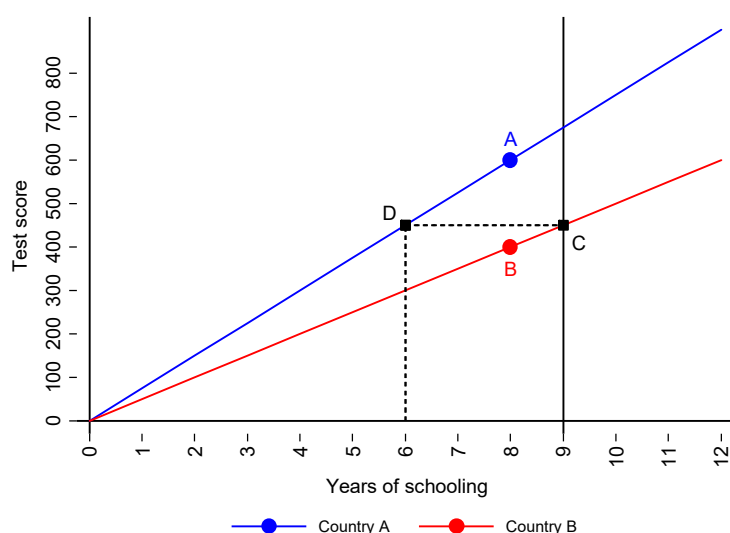
3.2 Years of schooling versus years of learning: when does learning begin?

The measure of relative learning productivity (defined in Equation 3) depends not only on the numerator, but also on the denominator (years of learning preceding the test). In the example above, we equate years of learning to years of schooling, but of course learning does not wait until children begin primary school. Every child acquires some language, mathematical concepts, reasoning skills, and socioemotional skills before arriving at school, and some systems may be better than others at fostering that acquisition. One indicator of this comes from the Multiple Indicator Cluster Surveys (MICS), which include assessments of whether young children between the ages of 3 and 6 can recognize 10 letters of the alphabet and whether they recognize the numbers 1 to 10.[7] While clearly not a complete measure of pre-school "learning," the results suggest that even as early as age 3—three years before most of them will start school—some children are already beginning to acquire these basic academic skills, and that the

---

[7] http://mics.unicef.org

share increases with age (Figure 3.2).

Figure 3.2: Proportion of children who can recognize 10 letters of the alphabet and the numbers 1 to 10, by age in months and country
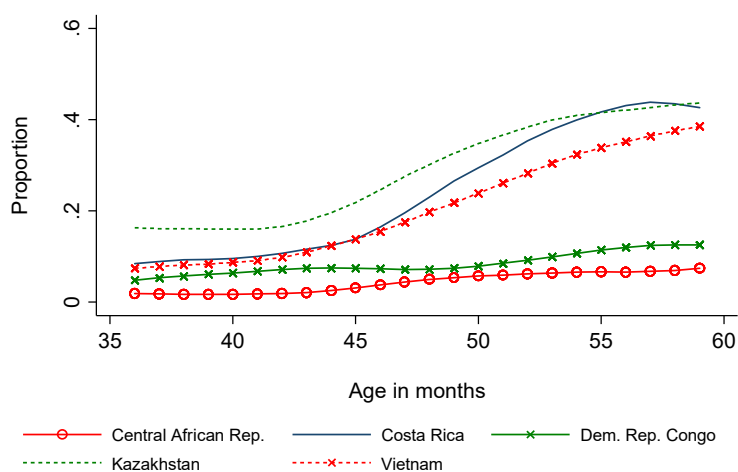
Source: Authors' analysis of Multiple Indicator Cluster Surveys (Round 4). http://mics.unicef.org

Accounting for the fact that years of learning may differ from years of schooling requires a modification to the *LAYS* calculation. The easiest way to show this is graphically. A key feature of the illustration in Figure 2 is that the ratio of test scores between any two points that are vertically aligned (vertical ratio of highest to lowest value of variable on the vertical axis) is equal to the ratio of years of schooling of any two points that are horizontally aligned (horizontal ratio of highest to lowest value of the variable on the horizontal axis). That is why the ratio of test scores at Grade 8 (400/600 = **2/3**) is the same ratio that is used to adjust the average years of schooling (9 × **2/3** = 6).

Figure 3.3 illustrates how this calculation needs to be modified if we assume that learning starts either "at birth," which we assume is 6 years prior to Grade 0 (we call this Grade "-6" for convenience), or that learning starts 3 years prior to Grade 0 (Grade "-3"). In this case the vertical ratio between points A and B is no longer the same as the horizontal ratio between the

years of schooling corresponding to points C and D.[8]

Figure 3.3: Illustration of the implications for *LAYS* of changing the assumption about when learning starts

A: Learning starts at Grade "-6"	B: Learning starts at Grade "-3"
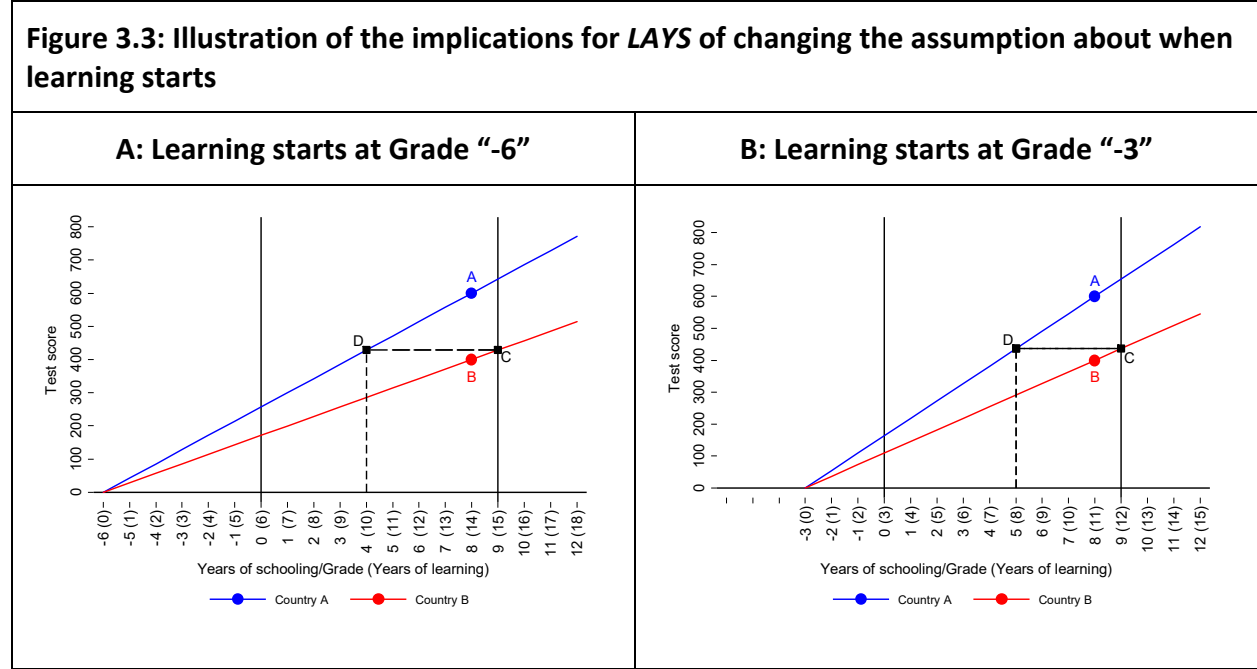
The change in assumptions leads to the following modifications to the relevant formulas. First, average learning per year (the slope of the average learning profiles) is now defined over the full number of years of learning, rather than just years of schooling. This means that Equation 3 is modified to:

$$L_c = \frac{T_c}{(Y_c + Y_c^p)} \tag{4}$$

where Y_c^p is the number of years of learning prior to Grade 0. Note that there is one additional assumption embedded in this framework: namely, that learning starts at the same point in different countries (that is, Y_c^p does not vary across countries: $Y_c^p = Y^p$).[9]

---

[8] The vertical ratio would be equal to the horizontal ratio if we were to use the "years of *learning*" scale for the latter. But since our ultimate objective is to derive a transformation for years of *schooling*, we carry out the modifications as described here.

[9] Because the learning assessment is typically done in a given grade across all countries—as with TIMSS in most countries— Y_c is also fixed, so the ratio of L_c to L_n will generally be the same regardless of when learning begins. We say this is "typically" true because, in some cases, assessments are administered in different grades. For example, the 2015 TIMSS was administered to Grade 9 students in Botswana and South Africa, whereas it was administered to Grade 8 students elsewhere.

The second modification involves the formula for converting the average years of schooling in one country into the number of years it would take another country to reach the same level of learning. (In graphical terms, this problem is equivalent to finding the x-axis value of the point on the blue line, labelled as point D in Figure 3.3, that is on the same horizontal line as point C on the red line.) The modified formula—that is, the *LAYS* formula from Equation 1, modified for learning that takes place before Grade 0—is:[10]

$$LAYS_c = [S_c \times R_c^n] - [Y^p \times (1 - R_c^n)] \tag{5}$$

The first term on the right-hand side of this equation is the same as before. The second term is the modification. As mentioned above, R_c^n is less than 1 by construction (if the highest-scoring country is used as the numeraire), so the *LAYS* value will decrease as Y^p increases: the modification will be larger if we assume learning starts earlier. Also, since the modification is larger when $(1 - R_c^n)$ is larger, the *LAYS* value will decrease as R_c^n decreases; in other words, the modification will be larger for poorer-performing countries.[11]

To provide a sense of how this change affects the magnitudes of the *LAYS* adjustment, Figure 3.4 shows the adjustment for the three cases discussed above: learning starting at Grades 0, -3, and -6. The adjustment is applied to the same set of 35 countries for which the approach was illustrated in Figure 2.1. Assuming that learning starts at Grade -3 leads to a *LAYS* value that is on average 0.69 years smaller than the *LAYS* based on assuming learning starts at Grade 0, while assuming learning starts at -6 leads to a *LAYS* value that is on average 1.39 years smaller. For example, 25- to 29-year-olds in New Zealand have 11.2 years of schooling on average; the *LAYS* adjustment brings that down to 8.9 years under the assumption that learning starts at Grade 0, 8.3 years if learning starts at Grade -3, and 7.7 years if learning starts at Grade -6. While there is no guarantee that country ranks will be preserved in these transformations, in practice they virtually are, with Spearman rank correlations among the three measures exceeding 0.99. So the

---

[10] This is derived from the formulas for the equations of each line, which are $y_A = L_A \cdot x_A + L_A \cdot Y^p$ and $y_B = L_B \cdot x_B + L_B \cdot Y^p$. If $y_A = y_B$ then $L_A \cdot x_A + L_A \cdot Y^p = L_B \cdot x_B + L_B \cdot Y^p$; therefore $L_A \cdot x_A = L_B \cdot x_B - Y^p(L_A - L_B)$; or $x_A = {L_B}/{L_A} \cdot x_B - Y^p {(L_A - L_B)}/{L_A}$; or finally $x_A = R_B^A \cdot x_B - Y^p(1 - R_B^A)$.

[11] It is clear from Figure 3.3 that LAYS could potentially be less than zero, once we assume learning to start before Grade 0. To avoid this, we add the further restriction that if the calculation yields a value that is less than zero, LAYS is set equal to zero.

14

major difference is not ordinal but cardinal: once we assume that opportunities to learn start well before primary school, low-learning countries see their *LAYS* values drop much farther—in some cases, to less than 2 years of quality-adjusted schooling.

Figure 3.4: Average years of schooling of the cohort of 25- to 29-year-olds, unadjusted and adjusted for learning using the *LAYS* adjustment with different assumptions about when learning starts

Source: Authors' analysis of TIMSS 2015 and Barro-Lee data.
Notes: Numeraire country is Singapore. Correlation coefficients between the three *LAYS* measures exceed 0.99. Spearman rank correlations between the three LAYS measures exceed 0.99.

4. Robustness of *LAYS* to the data source used

An important aspect of the *LAYS* approach is that it depends on the metric used to measure relative learning. There are at least four ways in which this statement is true. First, the particular units of the assessment used matters. Consider, for example, a transformation of test scores that preserves the average score across countries but changes the standard deviation of country averages around that average. The "distance" between any country and the top

15

performer would now be different, R_c^n would also be different, and the *LAYS* adjustment would yield a different result. Second, the particular assessment matters. Country rankings across assessment systems (for example TIMSS or PISA) tend to be fairly consistent, but a country's actual score and the value of that score relative to the top performer R_c^n would be different, again resulting in a different value for the *LAYS* adjustment. Third, the subject used to adjust for learning matters. While countries that tend to perform well in mathematics also tend to perform well in reading or science, the relative scores are not identical—suggesting that which subject is used might matter for calculating *LAYS*. Fourth, the choice of a fixed country as numeraire may be problematic if the best-performing country changes over time. We explore the empirical implications of each of these issues in turn.

4.1 The units of the assessment

A potentially important drawback of using the value of the test score on TIMSS (or an alternative similar measure) is that the *LAYS* calculation will be dependent on the units of that test—meaning that arbitrary rescaling of those units could therefore change the estimate of *LAYS*. An alternative approach could be to use an absolute measure of learning achievement. Results from the TIMSS assessment, as well as other international assessments such as PISA, are often reported as the share of test-takers who have reached a particular benchmark. These benchmarks are typically set by expert assessment of what level of mastery test-takers have achieved at given thresholds. We could use this share as the value of T_c in Equations (3) and (4) and proceed as before. That is, we redefine the *LAYS* formulas to be:

$$LAYS_c^B = S_c \times R_c^{Bn} \tag{6}$$

where R_c^{Bn} remains the ratio of "average learning per year" in country c relative to country n,

$$R_c^{Bn} = \frac{L_c^B}{L_n^B} \tag{7}$$

but "learning" at the level of each country is defined as the share of test-takers who have reached the given benchmark:[12]

_____

[12] Note that this formula is for the case where learning begins at Grade 0. The implications of allowing for learning

16

$$L_c^B = \frac{(Share > Benchmark)_c}{Y_c} \tag{8}$$

In this setup, years of schooling are being adjusted by the number of years that it would take the numeraire country n (given the average learning profile, now also defined in terms of shares who reach the benchmark) to get the same share of their students to the benchmark level as country c does.

The advantage of this approach is that it is independent of the units in which the assessment is reported. Changing those units would lead to a concomitant change in the value set for the benchmark, and the share of test-takers above and below that benchmark would remain unchanged.

Figure 4.1 illustrates, again for the 35 countries TIMSS in Figure 2.1, how implementing *LAYS* using the share of Grade 8 students who reach the "low" benchmark on the TIMSS mathematics assessment compares to the *LAYS* using the average TIMSS score (shown in the dark dots).[13] Note that an implication of this approach is that it "counts" improvements only at the lower end of the distribution of learning. If a country were to improve average learning levels but this improvement were to come from, say, students in the middle of the distribution, then this would not increase its *LAYS* value.

Panel A shows how R_c^n (the ratio that adjusts years of schooling for learning) is affected by this change in metric. Points below the 45-degree line are countries where the change leads to increasing the amount by which years of schooling are changed by the *LAYS* adjustment. The *LAYS* adjustment is exacerbated in countries that are already doing poorly. Countries where learning is poor on average (and where in addition it is highly unequally distributed) have especially large adjustments under the low benchmark approach. For example, in Morocco R_c^{Bn} is 0.20 lower than R_c^n (which is 0.61), and in South Africa and Saudi Arabia it is 0.23 and 0.25

---

to start earlier, as described above, would be similar in this case.

[13] The TIMSS "low" benchmark is set at score of 400. Students who have reached this benchmark have some basic mathematical knowledge such as adding or subtracting whole numbers, recognizing familiar geometric shapes, and reading simple graphs and tables (Mullis and others 2016). Singapore remains the best performer on this measure, with 99 percent of students reaching the benchmark (the same as in the Republic of Korea). The median percentage of test-takers who reach the benchmark across the 35 countries is 85 percent, with the lowest performers at 34 percent (Saudi Arabia and South Africa).

lower (where R_c^n is 0.53 and 0.59 in those countries respectively).

Figure 4.1: *LAYS* based on share reaching a low benchmark instead of average score

A: R_c^n as measured by share reaching low benchmark versus based on average score	**B: *LAYS* based on share reaching a low benchmark instead of average score**

Source: Authors' analysis of TIMSS 2015 and Barro-Lee data.
Notes: Numeraire country is Singapore. Illustration for case where learning starts at Grade 0. Correlation coefficients between the two *LAYS* measures is 0.97. Spearman rank correlations between the two *LAYS* measures is 0.97.

Panel B shows graphically the magnitude of the adjustments to *LAYS*. The general pattern of results is largely consistent with those based on average test scores. Both the correlation coefficient and the rank coefficient between the two measures are high, at 0.97. However, this alternative benchmark-based version does affect the point estimates for individual countries, most notably by further reducing the *LAYS* values of countries that were already performing poorly.[14]

---

[14] A similar exercise carried out using the "intermediate" benchmark exacerbates the loss due to the adjustment for learning, and the magnitude of this additional loss tends to be greater for those countries where the *LAYS* adjustment

The fact that the units of the assessment of learning matter is not in and of itself a problem for the *LAYS* approach. It does mean, however, that *LAYS* should be thought of not just "average years of schooling measured in terms of the learning productivity of the top performer" but as "average years of schooling measured in terms of the productivity of the top performer *according to the metric used to determine that productivity*." But given that the alternative unit-free measure yields results in line with the basic *LAYS* results, in practical terms that distinction may not be as significant as it first seems.

4.2 The assessment used: Using PISA instead of TIMSS

TIMSS is not the only assessment that could be used to calculate *LAYS*. PISA is another international assessment with wide coverage: in 2015, the PISA assessment covered 72 countries and economies (35 of which are in the OECD). PISA is an assessment of 15-year-olds in secondary school. Since these students are not necessarily all in the same grade, here we take a slightly different approach to calculating *LAYS*, using age rather than grade when we define the average learning profile. So, for example, we now divide the score on the assessment by 9 (in the equivalent of Equation 3), since this is the number of years of learning that a student who started learning at age 6 would have acquired by age 15. For alternative calculations that allow for learning to start before age 6, we could then proceed as before (that is, as per Equation 4) and add in additional years of learning. Implementing *LAYS* in this way using PISA 2015 mathematics scores yields the *LAYS* estimates shown in Figure 4.2, which are generally speaking in line with the results that use only TIMSS.

---

was already large using either the average or the "low" benchmark for the adjustment. See Annex 1 Figure 1.

Figure 4.2: *LAYS* using PISA assessment of 15-year-olds
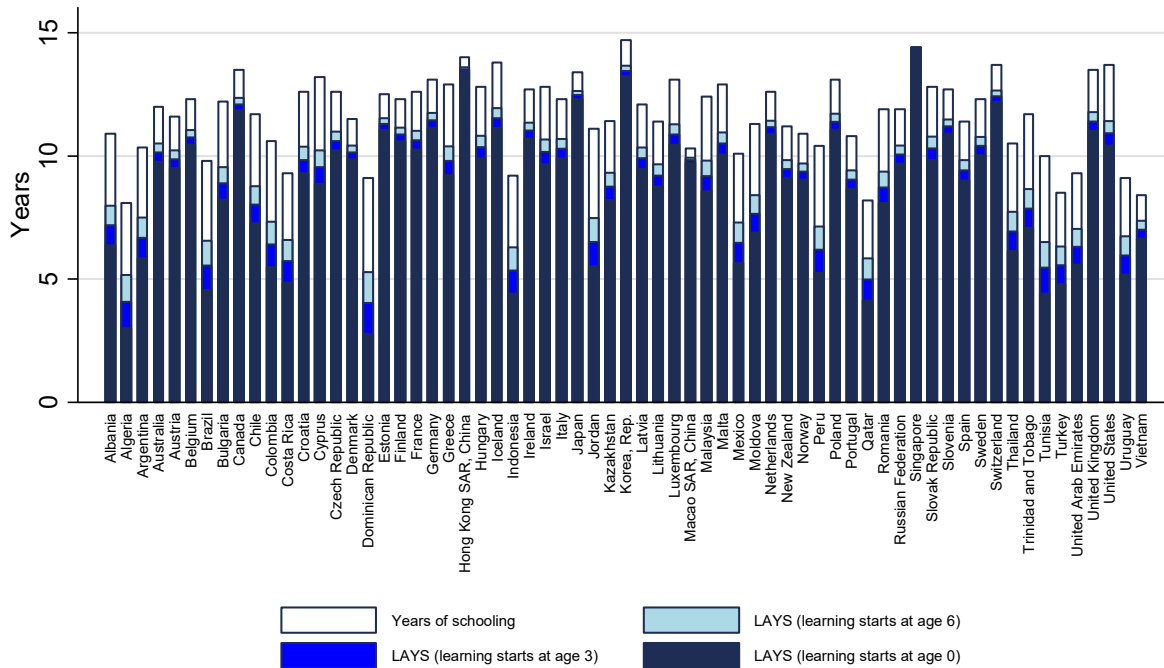
Legend:
- Years of schooling
- LAYS (learning starts at age 3)
- LAYS (learning starts at age 6)
- LAYS (learning starts at age 0)

Source: Authors' analysis of PISA 2015 and Barro-Lee data.
Notes: Numeraire country is Singapore. Illustration for case where learning starts at Age 6. Correlation coefficients between the three *LAYS* measures exceed 0.99. Spearman rank correlations between the three LAYS measures exceed 0.99.

There are 26 countries and economies that participated in the 2015 rounds of both TIMSS and PISA. For these countries we can calculate *LAYS* using both approaches and compare the estimates. We do not expect major differences across the approaches since the scores are similar; the mean TIMSS mathematics score for these 26 countries is 505 (standard deviation 55.7), the mean PISA mathematics score is 477 (standard deviation 45.6), and the correlation between the two is 0.91.[15]

---

[15] The fact that the average scores, and their standard deviations, are similar is not surprising. Both assessments were originally normalized to have mean 500 and standard deviation 100. For TIMSS, normalization was originally done for the (mostly high-income) countries that participated in the 1995 assessment. For PISA's math assessment, the normalization was done over the OECD countries that participated in 2003 (for reading, the year used was 2000).

The impact on *LAYS* of using one of these assessment systems versus another is illustrated in Figure 4.3. The *LAYS* results (right panel) are highly consistent: the correlation between the two *LAYS* measures is 0.98, and the rank correlation is 0.96.[16] Moreover, as the left panel shows, the tight relationship is driven by the strong correlation between the PISA and TIMSS measures of learning per year, and not solely by the fact that we are using the same measure of average years of schooling to calculate *LAYS* for both countries.

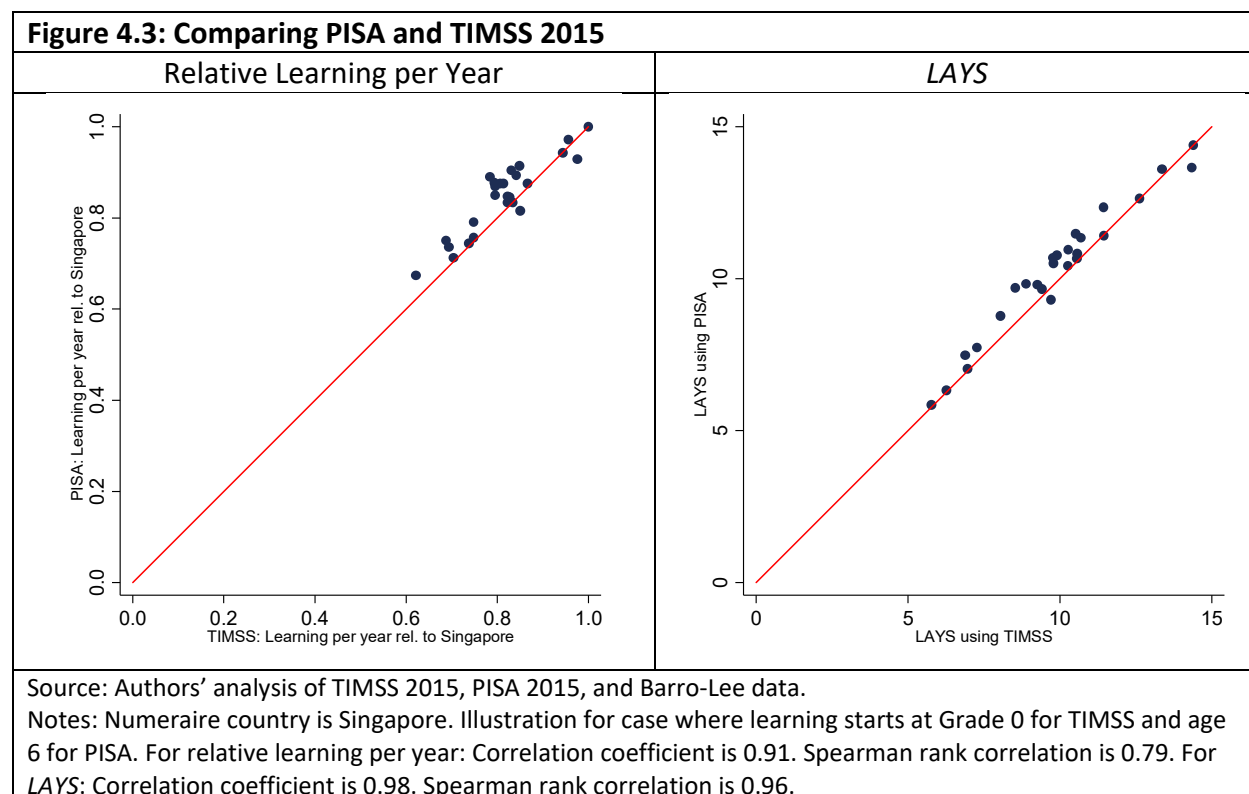**Figure 4.3: Comparing PISA and TIMSS 2015**

Source: Authors' analysis of TIMSS 2015, PISA 2015, and Barro-Lee data.
Notes: Numeraire country is Singapore. Illustration for case where learning starts at Grade 0 for TIMSS and age 6 for PISA. For relative learning per year: Correlation coefficient is 0.91. Spearman rank correlation is 0.79. For *LAYS*: Correlation coefficient is 0.98. Spearman rank correlation is 0.96.

4.3 The subject used to derive LAYS: Mathematics, Science, and Reading

The calculations above, whether based on TIMSS 2015 or PISA 2015, all use the assessment's mathematics score to derive the *LAYS* adjustment. It is natural, therefore, to ask whether *LAYS* estimates are sensitive to the subject used to quantify average learning-per-year. Each of these international assessments covers multiple subjects: TIMSS assesses Math and

---

[16] See Annex 1 for the implications of using the share who reach "Level 1" on the PISA scale.

Science, and PISA assesses Math, Reading, and Science. The cross-country correlation between these scores is high—that is, in countries where students do well in one subject, they tend to do well in other subjects too—so we would not expect using one or the other to lead to very different *LAYS* estimates. This is demonstrated in Figure 4.4: all points are close to the 45-degree line, meaning that the subject chosen does not make an appreciable difference to the estimate of *LAYS*. (In all cases, the correlation and Spearman correlations between these various measures of *LAYS* is above 0.98.)

Figure 4.4: *LAYS* using different subjects

LAYS based on TIMSS 2015 Science scores versus TIMSS 2015 Math scores	*LAYS* based on PISA 2015 Reading or Science scores versus PISA 2015 Math scores

Source: Authors' analysis of TIMSS 2015, PISA 2015, and Barro-Lee data.
Notes: Numeraire country is Singapore. Illustration for case where learning starts at Grade 0 for TIMSS and age 6 for PISA. For TIMSS: correlation coefficient is 0.99; Spearman rank correlation is 0.99. For PISA: between mathematics and reading: correlation coefficient is 0.99; Spearman rank correlation is 0.98; between mathematics and science: correlation coefficient is 0.99; Spearman rank correlation is 0.99

4.4 The choice of numeraire

Singapore is the highest performer on both the TIMSS and PISA mathematics assessments, which makes choosing it a logical choice for the numeraire in the illustrations of

LAYS above. A different approach could be to choose a basket of top-performing countries and adjust learning relative to the mean performance in that basket, as follows:

$$R_c^{Tn} = \frac{L_c}{\sum_{n=1,T}(L_n/n)} \tag{2}$$

The intuition of the *LAYS* adjustment is the same as before. The main difference is that for some of the countries in the basket (those whose value of L is above the mean for the top performers), by construction the *LAYS* estimate is now **greater than** their average years of schooling (since for them $R_c^{Tn} \geq 1$).

Figure 4.5 compares *LAYS* values using these two different numeraires—the mean of the top 5 performers (in terms of learning per year) versus the score of the top performer (Singapore)—based on both the TIMSS (left panel) and PISA (right panel) mathematics scores. Figure 4.6 shows how average years of schooling compares to *LAYS* in the two cases, again for TIMSS (top panel) and PISA (bottom panel). Since learning per year averaged across the top 5 performers will be less than that among the top performer, using the former of course yields *LAYS* estimates that are larger. Nevertheless, as Figures 4.5 and 4.6 show, the impact of changing the numeraire in this way is very small.

Figure 4.5: *LAYS* using top 5 performers versus top 1 performer (Singapore) as the numeraire

Source: Authors' analysis of TIMSS 2015, PISA 2015, and Barro-Lee data.
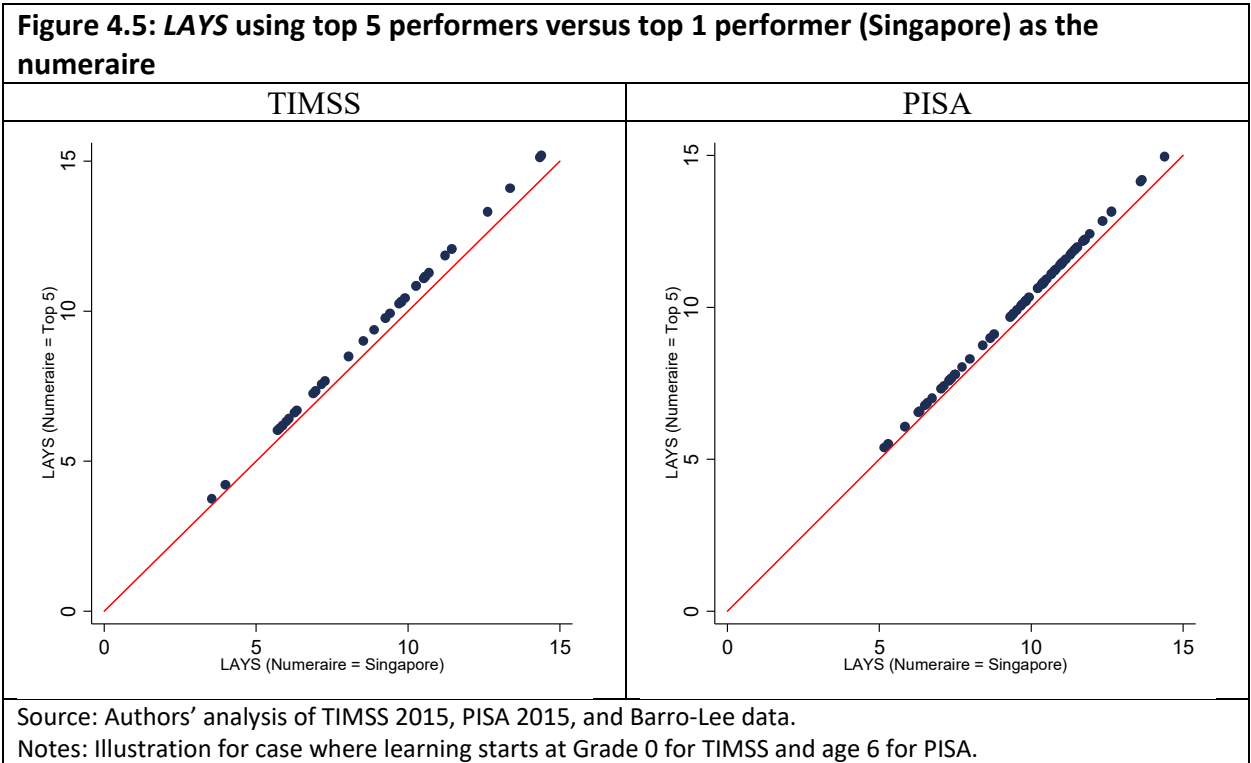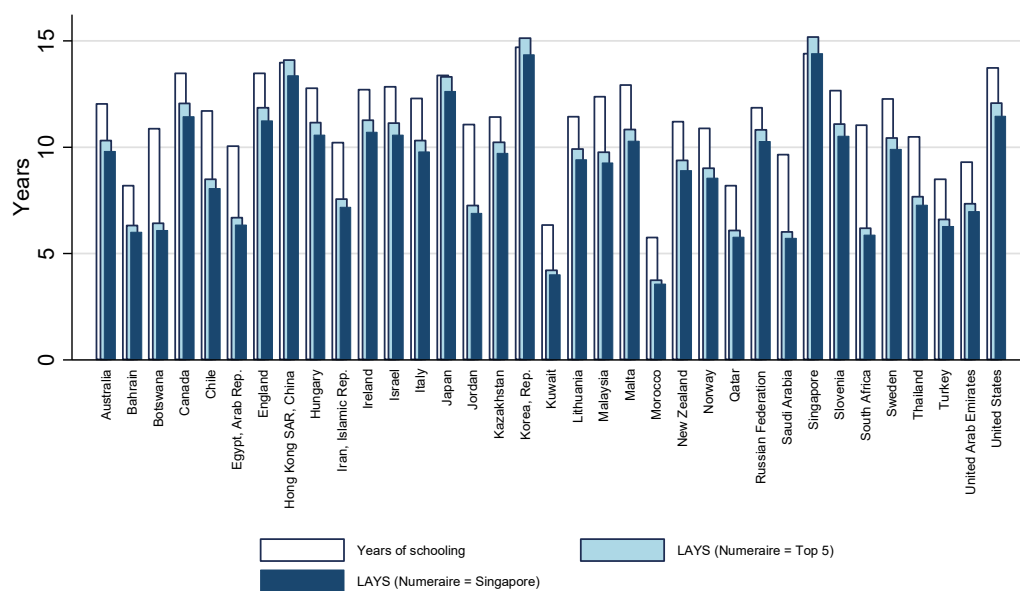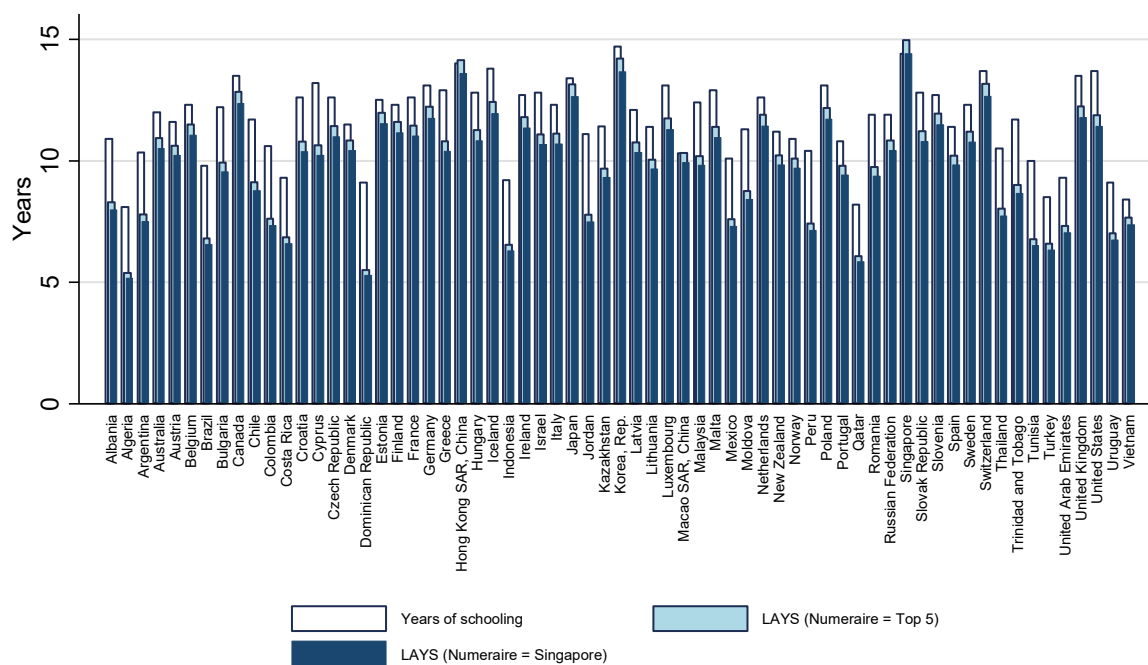Notes: Illustration for case where learning starts at Grade 0 for TIMSS and age 6 for PISA.

Figure 4.6: How the *LAYS* measure is affected by using the top-5 performers as the numeraire

Source: Authors' analysis of TIMSS 2015, PISA 2015, and Barro and Lee (2013) data.
Notes: Illustration for case where learning starts at Grade 0 for TIMSS and age 6 for PISA.

An alternative choice of numeraire could also be an "artificial" high performer. The advantage of this approach is that it is stable across subjects and over time by construction. For example, if we were to anchor *LAYS* to a Singapore in 2015 and another country, say the Republic of Korea, outperforms Singapore in the next round, we would then have to convert *LAYS* from Singapore-equivalent years to Korean-equivalent years, making a comparison difficult. In contrast, an artificial top performer with a fixed score would continue to serve as the anchor in this scenario. One reasonable performance benchmark would be the "advanced" benchmark of 625 on TIMSS and PIRLS, which is constant across subjects, grade levels and assessment rounds.[17]

5. Consistency of *LAYS* with other evidence

On the surface, the *LAYS* measure has some plausibility. It makes sense that we should value education systems (and, more broadly, societies) differently based on the amount of learning they deliver. And as we have seen, cross-country comparisons using the standard *LAYS* approach that use mean test scores from international assessments—which are admittedly based on a somewhat arbitrary scale—are similar to those based on the share of students reaching a particular level of learning proficiency (which are not scale-dependent).

This section evaluates whether other evidence is consistent with the assumptions and implications of the *LAYS* approach. Specifically, it explores: (1) whether students' learning gains across years exhibit local linearity, as assumed in the *LAYS* calculations; (2) whether observed returns to schooling are consistent with the quality adjustments implied by *LAYS*; (3) whether the *LAYS* learning adjustments are consistent with other test-score-based quality adjustments in the literature; and (4) how the findings from the *LAYS* approach, which relies on a multiplicative combination of quantity and quality, would compare with those of a linear approach that has been used on subnational data. The section concludes by comparing what these different approaches—existing and potential—would imply for the size of cross-country human capital gaps.

---

[17] The value of learning achievement in the numeraire country L_n (the Grade 8 average mathematics score for Singapore) used in Figures 2.1 and 3.4 is 621, so replacing that by 625 would barely change the results presented.

5.1 Local linearity of learning gains

In the initial example presented in this paper, *LAYS* is calculated using scores of 8th-graders. When we apply the relative learning in 8th grade to different average numbers of years of schooling, ranging from about 6 to 14 years of schooling, we are implicitly assuming that this learning ratio across countries remains constant—that each year of schooling is worth the same amount, in terms of learning, in any given country over that range (even as the learning rate differs across countries). While this assumption will not literally be true, we can test to see whether it makes sense as a rough estimate of cross-country differences. As the following sections will show, the evidence suggests that it does.

ASER data from India

One way to investigate learning trajectories is through an assessment that tests the same content across multiple grades. Tests such as PISA and TIMSS are tailored to each grade and age in which they are administered and are normed at the relevant level. This approach might produce linearity by test construction or scaling, even if the underlying learning trajectory on a constant measure would not be linear. To get around this problem, we analyze learning data from India collected for the Annual Status of Education Report (ASER), for which the NGO Pratham administers the same exact test to students from ages 5 to 16 across Grades 1 to 12 (ASER Centre 2017). The ASER data enable us to assess the rate of learning with a stable, comparable metric across grades and over time. To allow us to map out the specific trajectory for learning in school, we restrict our sample to school-going children.[18] Figure 5.1 shows that students learn to divide along an S-shaped learning trajectory, with a locally linear interval from Grades 6 to 10. Figure 5.2 shows how often this locally linear interval appears at the subnational level, using 2012 ASER data for 31 Indian states. We compare observed learning trajectories to a projected linear trend and find remarkable alignment, indicating local linearity across most states in that year. We see this trend repeated across nearly all states and all years from 2008 to 2012 (see Annex 1, Figure 3 for all results).

---

[18] Note that this is comparing different cohorts of students at different grades, not the same students over time.

Figure 5.1: Learning Trajectories in India

A: Percent who can do division, by grade
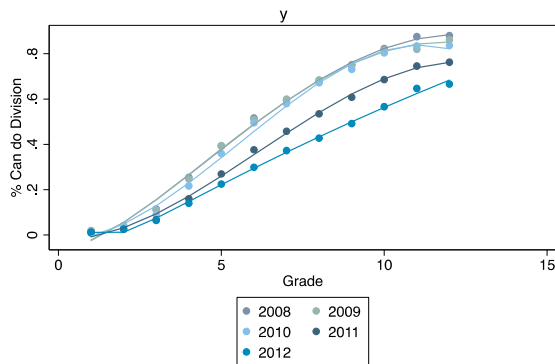
B: Percent who can do division, by grade *(linear trend between Grades 6 and 10)*
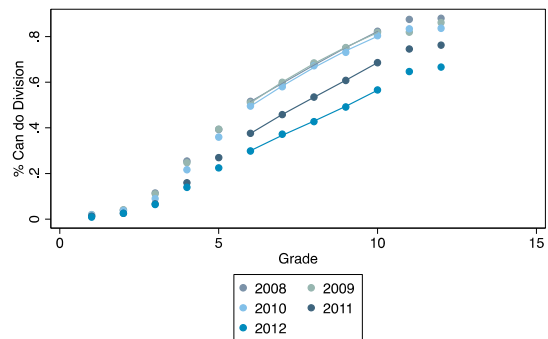
Source: Authors' analysis of Indian ASER data (2008-2012)

Figure 5.2: Learning Trajectories in India, by State (2012)
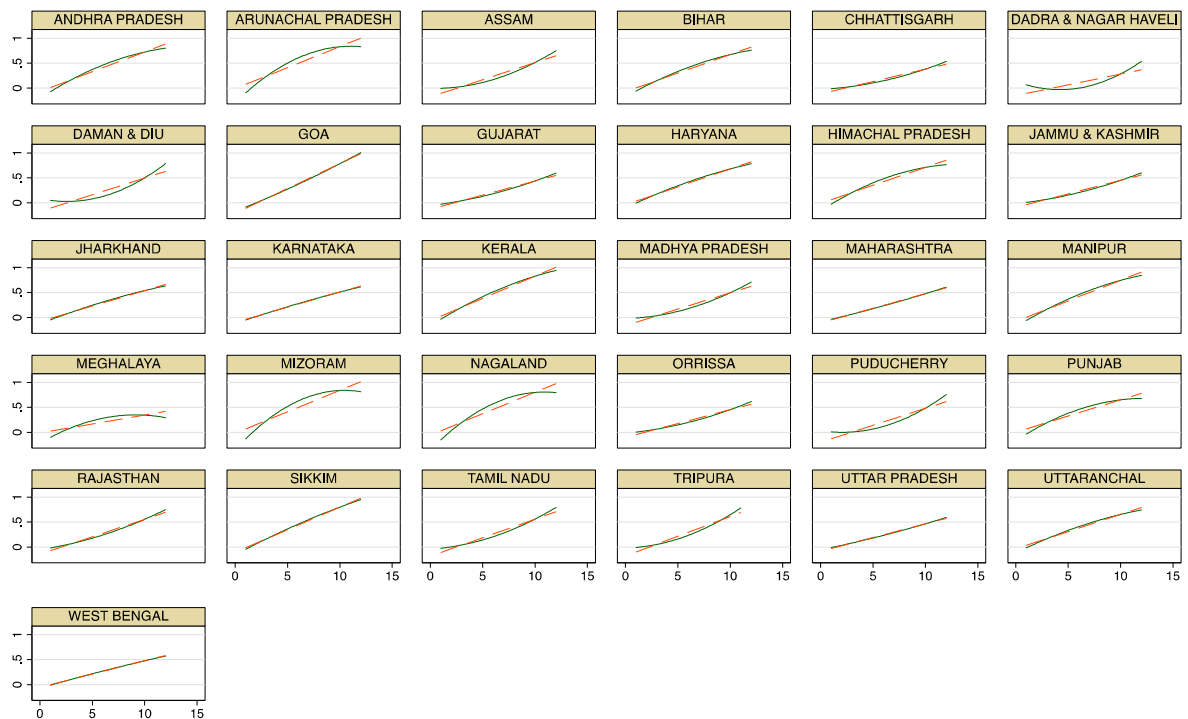
PISA data across grades

Another way to test the local linearity assumption is through inter-grade comparisons of

27

test-takers. PISA can be used for this purpose, leveraging the fact that the 15-year-olds who take PISA can be in different grades. We can therefore see whether the scores increase approximately linearly from one grade to the next.[19] Figure 5.3 shows, for each country, the mean math score for each grade in which at least 100 students took the test.[20] The dark line segments connect these averages across grades. The red (dashed) lines connect the scores in the lowest and highest grade—that is, they map out what a perfect linear trajectory would look like. In many countries the two lines are indeed quite close, indicating the linearity is a reasonable assumption in this range.

Through methods like this, the OECD has estimated the slope of the PISA learning curve. As a rough rule of thumb, it estimates that each 30- to 35-point gain on PISA is roughly equivalent to an additional year of education, on average across all countries—or in other words, each year of education is worth about 30 to 35 PISA points (see OECD 2013, 2016; also see the discussion in Appendix D of Jerrim and Shure 2016). Imagine that we can extrapolate this slope backwards over the length of a student's life, from the age at which the student takes PISA. (Of course, young children would all score zero on the actual PISA, making the slope of the measured learning curve horizontal at young ages; but assume that the slope represents performance on some age-appropriate measure of underlying cognitive skills.) Then given that PISA is a test of 15-year-olds, projecting backward using a 35-point-per-year slope from the PISA average score of 500 gives a score of zero around the age of zero, which is consistent with learning starting at right around birth. In other words, the data are consistent with learning that accumulates from birth to age 15 at a rate equal to that of the linear trajectory found in the observable age range.

---

[19] This is an imperfect test, because 15-year-olds are not randomly allocated across grades. As discussed below, those in a lower grade than the typical 15-year-old might have lower achievement because they have been held back; those in a higher grade might be there because they are high achievers. This is why we complement this test with another approach below that controls for selection.

[20] Only countries for which there are at least 100 students in each of at least 3 grades are included in this analysis.

Figure 5.3: PISA average math scores by grade (compared to linear trend)
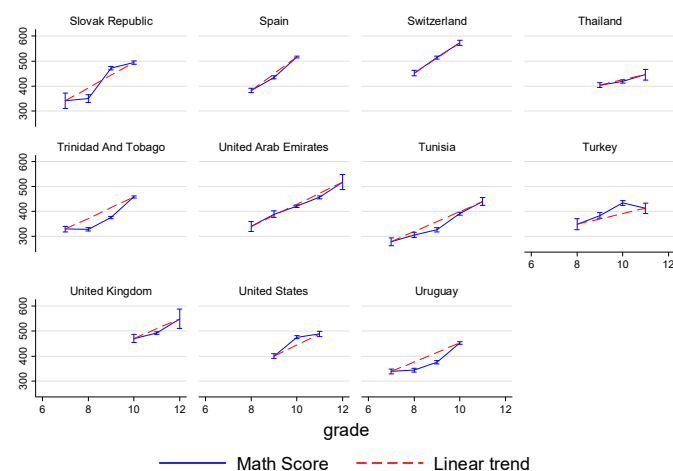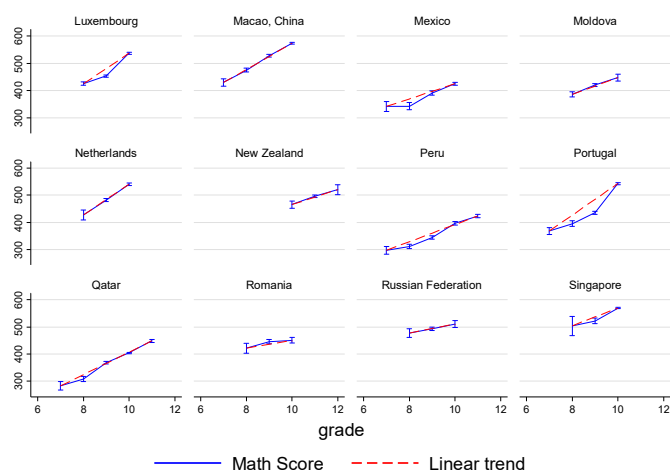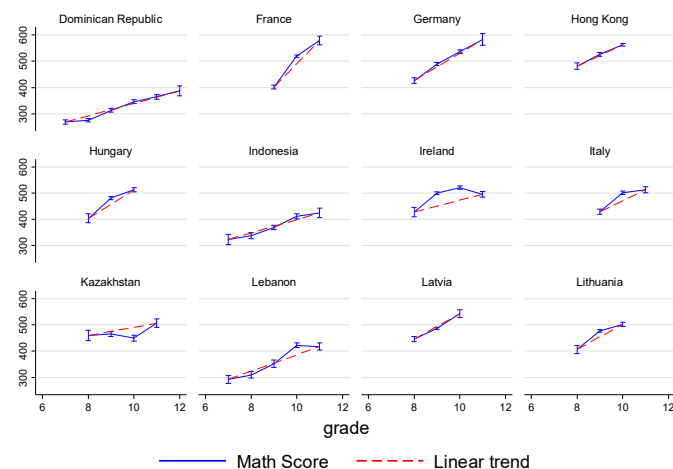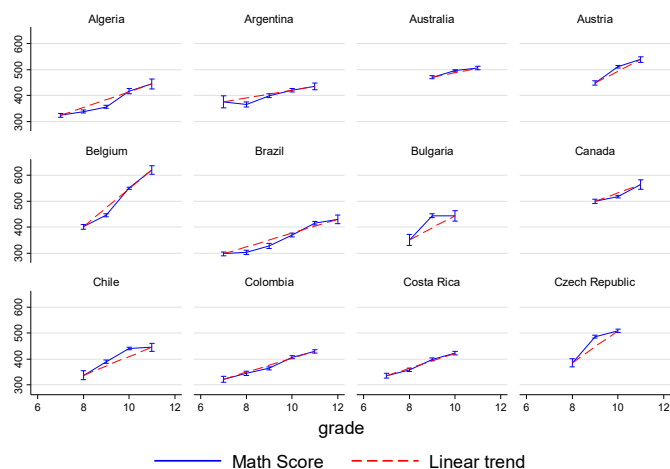
Source: Authors' analysis of PISA 2015 database. Vertical bars show the 95 percent confidence intervals around the calculated grade-specific test scores.

Accounting for sample selection

The above two analyses—using ASER or PISA data—suffer from the potential fact that the profiles of students might be different across grades. This could be, for example, because poorer-performing students drop out at higher grades, or because they are held back and are therefore more concentrated in lower grades. More generally, students who are a grade ahead (or behind) their peers might be different in other ways, and not just in the amount of schooling they have had. In other words, selection on unobservables may bias the results. We therefore now turn to approaches that account for potential selection.

The approaches we use deploy a combination of regression discontinuity and instrumental variables estimation. Specifically, we examine random variation in birth month across school entry cutoffs. Students randomly born a few days after the school cutoff enroll up to a year later than their peers who are just a few days older. This enables us to estimate the causal effect of an additional year of schooling.

There is an extensive literature using random variation in birth month across school-entry cutoffs to identify the effect of *relative age* differences on learning.[21] This literature relies on estimating test score variation across exogenously determined relative ages. Our approach is inspired by this literature and repurposed to assess *relative grade* rather than *age* since we care about how much is learned over the course of a year of schooling.[22] Specifically, we use PISA 2015 data, which assesses 15-year-olds across multiple grades. This dataset includes cohorts born in 1999 and 2000. Since PISA data only includes birth month and year, we are unable to examine precise effects around exact school entry dates as is typical in the literature. Instead, we explore effects around school entry months, which proxies for specific entry dates. An advantage of the PISA data is that it codes a variable for the gap between the grade a student is in and the expected grade given that country's school entry age laws. For example, in Finland where school starts at age 7, students are expected to be in grade 9 by age 15. In Mexico, students start at age 6 and are expected to be in grade 10 by age 15. This means a PISA-taking

---

[21] See, for example, Angrist and Krueger 1991; Dobkin and Ferriera 2010; Fredriksson and Ockert 2014; Smith 2009; Robertson 2011; Puhani and Weber 2007; Kawaguchi 2011; McEwan and Shapiro 2008; Bedard and Dhuey 2006; Elder and Lubotsky 2009; Pehkonen et al. 2015; and Crawford et al. 2010.
[22] This is similar to work being conducted by Singh (2017) using data from the Young Lives project.

student in grade 9 in Finland would be coded as 0, whereas in Mexico the same student would be coded as -1. Thus, this variable is sensitive to country-specific grade progression and is comparable in pooled or cross-country regressions.

Information on formal school entry date cutoffs, and on how well these cutoffs are enforced, is hard to locate and verify. Moreover, school entry dates often vary within country and over time. Given these difficulties, rather than using the legal cutoffs, as a first step to screening countries we examine discontinuities in the data in relative grade by birth month. We restrict our sample to countries with large samples on both sides of the cutoff. Four countries—Mexico, the Republic of Korea, Slovakia, and Thailand—meet this criterion. We find large and significant shocks to relative grade: being born just before the school entry cutoff (versus being born just after it) exogenously increases relative grade by 0.55 to 0.6 grades. Figure 5.4 shows graphs of how month of birth indeed affects relative grade in these four countries; and the month of the discontinuity corresponds to legal school entry month. For example, in Korea the school entry cutoff is January 1st, and we see a sharp discontinuity in relative grade between December and January. In Mexico and Slovakia, the school entry cutoff date is September 1st and we see corresponding discontinuities between August and September. It is worth noting that in Mexico the discontinuity occurs right at the start of August, whereas in Slovakia this starts to occur earlier. This suggests that while this cutoff is enforced in Slovakia, it might be less strict.

Figure 5.4: Relative grade by birth month

Source: Auth

31
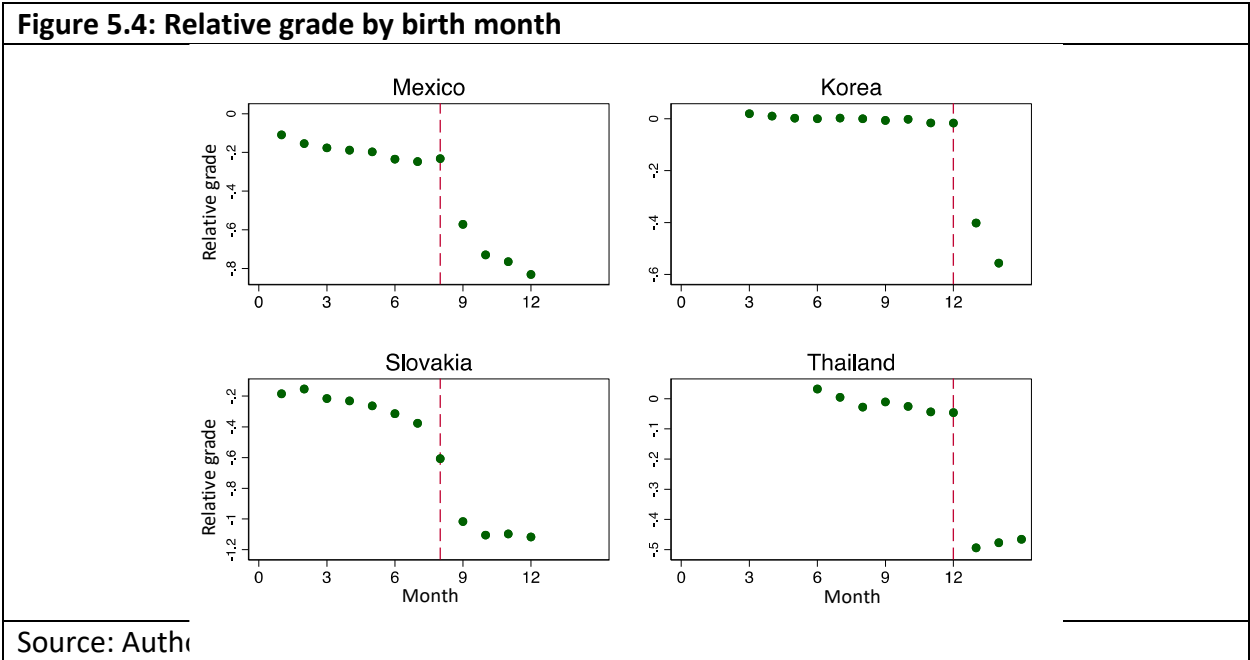
In these countries, we see a correspondingly large discontinuity in test scores. Figures 5.5 and 5.6 depict the results graphically and Table 5.1 quantifies them. The figures show a sharp break in test scores at the grade cutoff, both in the individual countries (for some subjects) and for the pooled sample. In the regressions, the OLS results show that the additional year of schooling around the cutoff is correlated with between 39 and 48 points in math and reading. When we control for selection using regression discontinuity and instrumental variable approaches, the coefficient shrinks as expected. Nevertheless, they suggest that an additional year of schooling increase scores by between 12 and 35 points, depending on the method and sample used.

Figure 5.5: Test Scores by Birth Month: Individual countries and subjects

Source: Authors' analysis of PISA 2015 data.
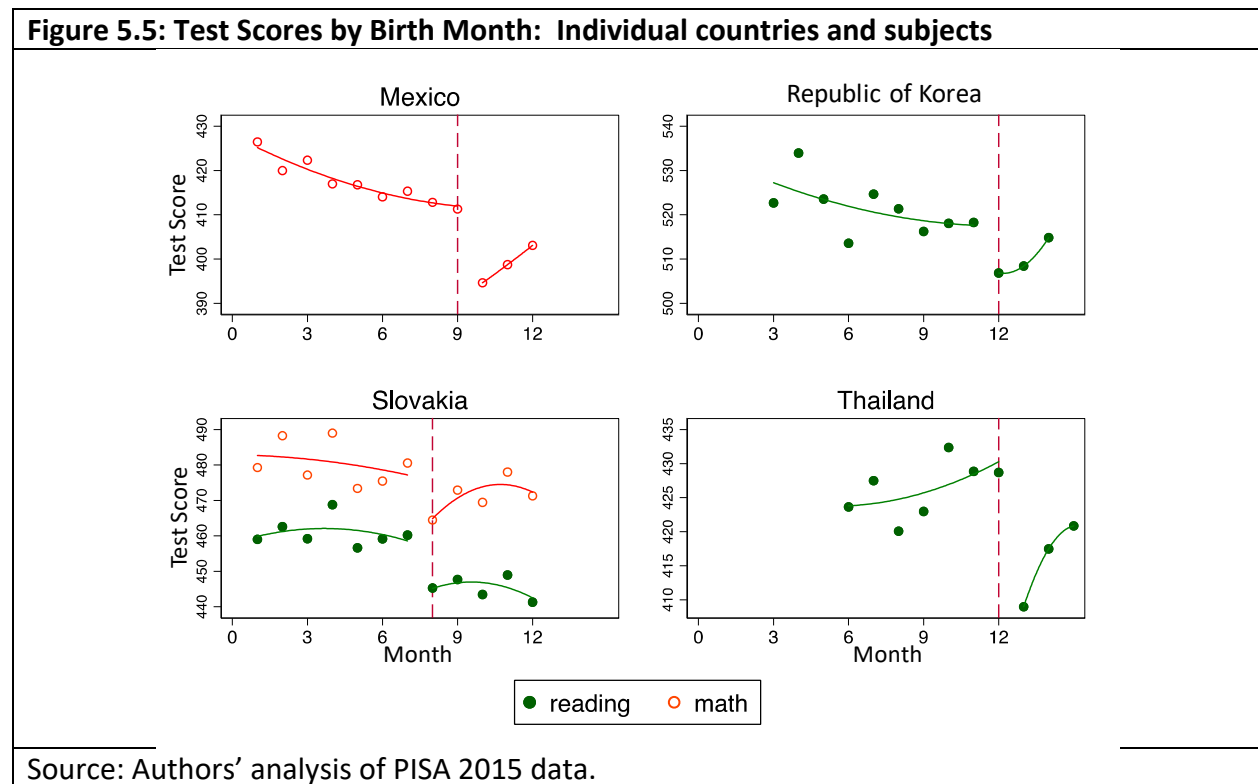
Figure 5.6: Test Scores by Birth Month: Pooled Data (Mexico, Republic of Korea, Slovakia, Thailand)
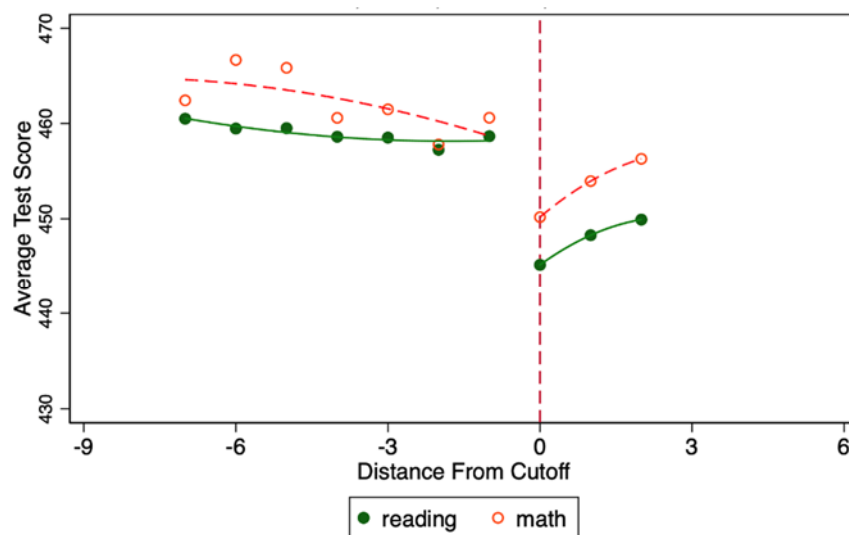
Source: Authors' analysis of PISA 2015 data.

Thus, even after accounting for potential selection effects, we reach a similar conclusion as in the analysis in the previous section. Given the year-to-year rate of student learning observed in the data, if we project learning backwards at a constant rate using the larger estimated coefficients (31 to 35 points, from the 2SLS estimates in the columns 3 and 7), this is consistent with learning starting at birth and continuing at (on average) the same rate observed in the data. If instead we project backwards using the smaller coefficient estimates (12 to 20 points per year), this would imply that students start life with some initial endowment of cognitive ability.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reading				Math			
	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
Grade	47.48***		35.16***	16.66***	38.89***		31.82***	12.37***
	(0.992)		(2.227)	(3.309)	(1.014)		(2.270)	(3.373)
Assigned Relative Grade		18.31***				20.23***		
		(1.336)				(1.326)		
Observations	26,220	26,220	26,220	12,917	26,220	26,220	26,220	12,917
R-squared	0.080	0.007	0.075	0.046	0.053	0.009	0.051	0.025

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

same way, but include only the sub-sample around the discontinuity cutoff.

While each of the analyses presented in this subsection has potential drawbacks, together they suggest that learning trajectories have a plausibly local linear trend, especially across the grades that we are interested in, and that this is true for multiple specifications and tests. Moreover, on balance the results suggest that the calculated slopes are consistent with a learning trajectory that starts at birth rather than at age 6—meaning that the *LAYS* adjustment ratio is really a measure of learning productivity of a *society*, and not just its schools.

5.2 Evidence on schooling quality derived from labor market returns

Second, we examine whether labor-market evidence is consistent with the LAYS assumption and implications. As noted above, the *LAYS* measure shows that the stock of schooling—once adjusted for quality of learning—in some countries is much lower than the standard measures indicate. Of course, this is true not just for societies, but also for the individuals in those societies. But this throws a wrench into the standard calculations of returns to education. The private Mincerian return to schooling is typically calculated to be between 8

and 10 percent per year of schooling.[23] But if the average individual now has far fewer years of effective schooling, this would imply that the returns to effective schooling (or learning-adjusted schooling) are much higher—up to 16 to 20 percent per year. Is this plausible?

It isn't possible to test this directly, because we don't directly observe the returns to effective years of schooling. But what we can do is flip that around: test whether differences in the implied quality of schooling explain differential returns to schooling within a common labor market. For this purpose, we can draw on Schoellman (2012), who calculates the returns to schooling in the US market for immigrants from many countries. These data have the advantage of holding constant national labor-market conditions, and thus the demand for schooling, so that the returns reflect only the quality of schooling and not demand-side differences. He finds that an additional year of Mexican or Nepalese education raises the wages of Mexican or Nepalese immigrants by less than 2 percent, while an additional year of Swedish or Japanese education raises the wages of Swedish or Japanese immigrants by more than 10 percent.

The left panel of Figure 5.7 compares Schoellman's estimates of returns to the learning adjustment ratio underlying *LAYS*, following an example from Figure 1(b) in his paper. It shows a strongly positive relationship between the two measures: countries whose students do better on learning metrics also have emigrants who earn higher returns for each year of schooling in the US labor market. There are a few significant outliers: most notably, the Republic of Korea (represented by the bottom-right point) does much better on learning than on the returns measure, while on the other side, South Africans (left-most point) do much better on labor-market returns than on learning. But on the whole, the figures and the statistical correlations strongly suggest that the relative learning measure is picking up something economically meaningful in terms of skills.

---

[23] Psacharopoulos and Patrinos (2018), for example, find that the global average is 9 percent per year of schooling.

Figure 5.7: Comparisons between LAYS and Schoellman (2012)'s returns-based measures

Returns to schooling in the US labor market for immigrants from different countries (Schoellman 2012) versus learning-per-year relative to Singapore	Human capital derived using Schoellman's approach applied to Barro-Lee data versus LAYS

Source: Authors' calculations, using US labor-market returns data from Schoellman (2012) and *LAYS* analysis using TIMSS.

Notes: For ln(h), see text for details. For LAYS, numeraire country is Singapore. Illustration for case where learning starts at Grade 0 for TIMSS. For relative learning per year: Correlation coefficient is 0.23 and Spearman rank correlation is 0.29. Excluding just three countries (Kuwait, South Africa, Rep. of Korea) increases these correlations to 0.65 and 0.64, respectively. For LAYS: Correlation coefficient is 0.61 and Spearman rank correlation is 0.68. Excluding the same three countries increases these to 0.84 and 0.85, respectively.

Schoellman (2012) also derives a measure of the stock of human capital in a country, built on his measure of quality-adjusted years of schooling, and based on the following equation:

$$\ln(h_c) = \frac{S_c Q_c}{\eta}$$

where S_c is years of schooling in country c, and Q_c is schooling quality in country c and is implemented as the wage increment to an additional year of schooling of those educated in that country while working in the US. The parameter η is included to account for the fact that years of schooling acquired themselves might be higher because of school quality; based on his empirical estimates, Schoellman sets this parameter equal to 0.5. To compare our *LAYS* approach to that of Schoellman (2012), we calculate his estimate of $\ln(h_c)$ using Barro-Lee

36

estimates of years of schooling of the cohort of 25- to 29-year-olds for S_c, Schoellman's estimates of labor market returns for Q_c, and his preferred value of 0.5 for η. The right panel of Figure 5.7 shows that there is broad agreement between the two measures (excluding just three outliers, the correlation coefficient between the two measures is 0.84, and the rank correlation is 0.85).

5.3 Quality adjustments based on grade-to-grade test-scores increases
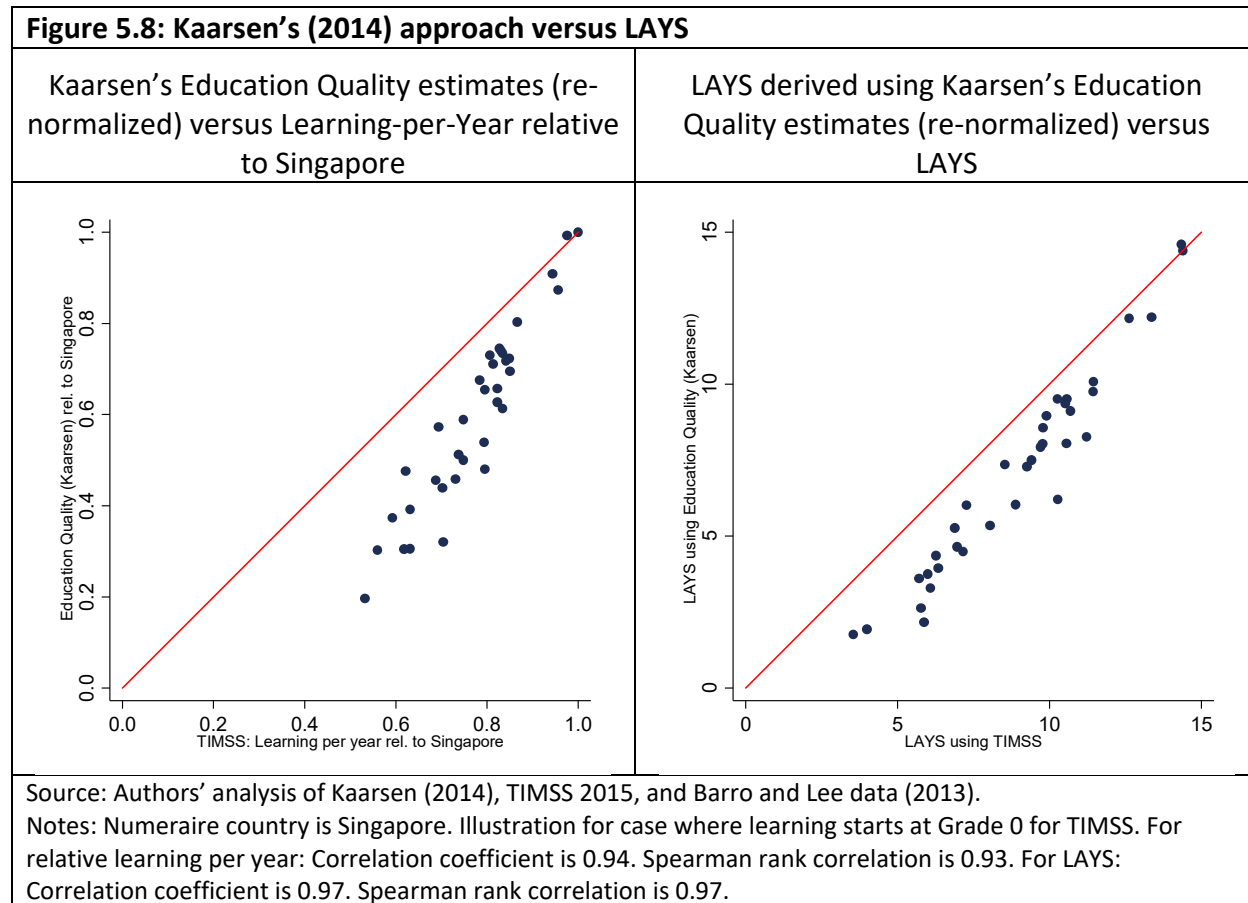
Third, we can test the plausibility of *LAYS* by comparing its findings with those of other efforts to adjust schooling for quality. Beyond *LAYS* and Schoellman's returns-based approach, there are alternative ways to adjust years of schooling for quality to allow cross-country comparison. One relevant recent study that (like *LAYS*) relies on measured test scores to do this is Kaarsens (2014), which derives measures of education quality across countries using multiple rounds of TIMSS (1995-2011). The analysis exploits the fact that in 1995, students were tested in Grades 3 and 4 at the primary level and at Grades 7 and 8 at the secondary level.[24] It uses these data to calibrate a model that simultaneously estimates overall average annual test score growth (that is, the average changes from Grades 3 to 4, and from 7 to 8), along with "education quality," which is (essentially) a country-specific fixed effect calculated after controlling for the overall average test-score growth.[25] The approach then normalizes the quality measure relative to the United States. To more directly compare the approach to the *LAYS* approach, we renormalize the "education quality" estimates so they are expressed relative to Singapore's; we then recalculate *LAYS* by multiplying that relative quality measure by the Barro-Lee average years of schooling of the cohort of 25- to 29-year-olds.

The left panel of Figure 5.8 compares Kaarsen's (renormalized) education quality variable to the measure of learning per year relative to Singapore that we use (that is, the data underpinning Figure 2.1); the right panel compares the *LAYS* estimates that follow from the two approaches. Kaarsen's approach suggests that there are larger differences across countries in

---

[24] The data for Grades 3 and 7 were made available to him as a result of personal communication with the IEA, the organization that produces TIMSS.
[25] According to Kaarsen's calculations, the average TIMSS test-score gain in math at the secondary level is 29.4 points, and in science at the secondary level it is 36.5 points.

quality than our approach does—or, in other words, that our approach is more conservative in discounting the schooling of poorer-performing countries. Nevertheless, the *LAYS* estimates derived under the two approaches are highly correlated.

Figure 5.8: Kaarsen's (2014) approach versus LAYS

Kaarsen's Education Quality estimates (re-normalized) versus Learning-per-Year relative to Singapore	LAYS derived using Kaarsen's Education Quality estimates (re-normalized) versus LAYS

Source: Authors' analysis of Kaarsen (2014), TIMSS 2015, and Barro and Lee data (2013).
Notes: Numeraire country is Singapore. Illustration for case where learning starts at Grade 0 for TIMSS. For relative learning per year: Correlation coefficient is 0.94. Spearman rank correlation is 0.93. For LAYS: Correlation coefficient is 0.97. Spearman rank correlation is 0.97.

5.4 Linear versus multiplicative combinations of schooling quantity and quality

A final consistency check compares LAYS with estimates that combine schooling quality and quantity in a different way. As explained in Section 2, the LAYS measure, like Schoellman (2012) and Kaarsen (2014), uses a multiplicative approach to deriving quality-adjusted years of schooling. In essence, all three use a variant of the form:

$$(Adjusted\_Y)_c = Y_c \times q_c$$

In our case, $Adjusted\_Y$ is *LAYS, and* q_c is equal to R_c^n, which represents test scores relative to those in a numeraire country. Hanushek, Ruhose, and Woessmann (2017) take a different

approach in their analysis of the impact of "knowledge capital" on GDP growth across US states. As a first step, they define quality-adjusted human capital as:

$$h = e^{rS+wT}$$

where S is years of schooling and T is a measure of test scores.[26] They draw their estimates of r and w from the micro-economic literature that estimates hourly earnings as a function of S and T in samples of the working-age population (taking care to use estimates of r only from models that have also controlled for T, and estimates of w only from models that have controlled for S). Their preferred estimates are $r = 0.08$ and $w = 0.17$.[27] They then use a development accounting framework to estimate the share of cross-state GDP growth that is attributable to knowledge capital (and within that share, how much is due to years of schooling and how much to test scores).

It is not straightforward to compare their measures to ours, since Hanushek, Ruhose, and Woessmann (2017) calculate their estimates for US states, not for countries. Nevertheless, to get a rough sense of how the measures compare, we implement their approach to characterizing h on the same Barro-Lee years of schooling estimates and TIMSS test scores used to derive Figure 2.1, and compare it to our measure of *LAYS* derived from that sample.[28] Figure 5.9 illustrates how the measures derived from these two approaches compare across this set of countries. Clearly there is a high degree of correspondence between the two, with a correlation (and rank correlation) coefficient exceeding 0.9.

---

[26] While, of course, h here is a nonlinear function of S and T, the model actually uses ln(h), which is a linear combination of the two. In the calculation of h for each State, the authors are very careful about accounting for the effects of migration (both domestic and international) on the estimates of T.

[27] Prior to applying these weights, they normalize test scores to have a standard deviation of 1.

[28] Their estimate of test scores has a mean of 5 and a standard deviation of 1 across the US sample, so is comparable to that in TIMSS (if we divide scores by 100). We do not carry out the careful adjustments for migration that Hanushek, Ruhose, and Woessmann (2015), which is, in part, why we refer to this as a rough comparison.

Figure 5.9: Hanushek, Ruhose, and Woessman (2017)'s approach versus LAYS

Human capital derived using Hanushek, Ruhose, and Woessman's approach applied to TIMSS and Barro-Lee data versus LAYS
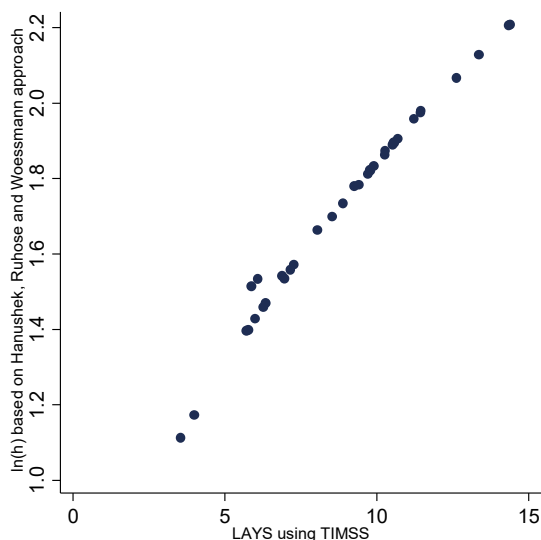
Source: Authors' analysis of TIMSS 2015, and Barro and Lee data (2013).
Notes: For ln(h), see text for details. For LAYS, the numeraire country is Singapore. Illustration for case where learning starts at Grade 0 for TIMSS. Correlation coefficient is 0.99. Spearman rank correlation is 0.99.

5.5 Quality adjustments using different approaches: How do they compare?

We have illustrated how each of the approaches described above compares to the *LAYS* approach by graphing one versus the other. But this doesn't provide a sense of how the different approaches would characterize international gaps in quality-adjusted years of schooling. To provide a sense of this, for each potential approach to measuring schooling, we calculate the average value for the four top-performing countries and the four bottom performers on that measure, and then take the ratio of these two values. The results are reported in Table 5.2. To understand how to interpret it, take the example of Row (1), which examines the simple quantity-based years-of-schooling measure. Row (1) shows that the average years of schooling of 25- to 29-year-olds in the four lowest-performing countries (in this sample) is 7.1, compared with 14.2 among the four highest-performing countries; the global "gap" as measured by the ratio of best to worst performers is therefore 2.0. If one were to just look at scores on the TIMSS 2015 Math assessment, the gap would be 1.6. Unsurprisingly, the approaches that combine quantity and

quality multiplicatively tend to show larger gaps; the *LAYS* approach is the most conservative of the three discussed here, yielding a top-to-bottom ratio of 2.9 versus 3.3 and 6.3 for the other approaches). The last approach, which is the additive approach from section 5.4, yields a smaller gap: its top-to-bottom ratio of 1.7 is similar to that of the simple TIMSS approach.

Table 5.2: How different approaches characterize the human capital gap between the bottom and top performers				
Approach	Based on	Average value for four bottom-performing countries	Average value for four top-performing countries	Ratio (top to bottom performers)
(1) Average years of schooling of 25- to 29-year-olds	Barro-Lee	7.1	14.2	2.0
(2) Math test scores on international assessment	TIMSS	377.5	601.8	1.6
(3) Learning-Adjusted Years of Schooling (LAYS)	See text	4.8	13.7	2.9
(4) "Human capital" based on returns to schooling of US immigrants	Schoellman (2012)	0.9	3	3.3
(5) Quality-adjusted years of schooling based on test score gains across years	Kaarsen (2017)	2.1	13.3	6.3
(6) "Human capital" based on linear combination of years of schooling and test scores	Hanusheck, Ruhose, and Woessman (2015)	1.3	2.2	1.7
Source: Authors' analysis based on Barro and Lee (2013), TIMSS 2015 data, Schoellman (2012), Kaarsen (2017), and Hanushek, Ruhose, and Woessman (2017). See text for details on how each measure is implemented. Note: The sample of countries is the same 35 countries as that for Figure 2.1, except in the case of Row (4). Under the Schoellman approach, data are not available for 8 of those countries, and the analysis underlying Row (4) therefore only includes 27 countries.				

These various approaches clearly have different implications for how one assesses

international gaps in quality-adjusted years of schooling, but they are also clearly related to one another and generally point toward wider gaps in quality-adjusted schooling than in the standard measure. Of these measures, *LAYS* is relatively conservative in the way it expands the gap compared to most other approaches, which we see as an advantage for its acceptance in policy circles. The next section discusses why we believe *LAYS* is an attractive measure to use for policy and research purposes, and how we should think about what *LAYS* doesn't include.[29]

6. Using *LAYS* to measure education progress

Using *LAYS* in place of the standard years-of-schooling macro metric of education could have various incentive effects on policymakers and educators. This section explores those implications—both the positive effects and possible unintended negative incentives.

6.1 Advantages of LAYS for research and policy

As discussed in the opening section of this paper, *LAYS* is intended to provide a more accurate reflection of educational progress than the standard summary measures. A major reason for doing this is to improve our understanding of how education affects, or is affected by, other important development outcomes like economic growth, inequality, governance, or health. If (measurable) learning is what mediates the effects of schooling, then omitting it from the analysis will cloud our understanding of those relationships. An example is Hanushek and Woessmann's finding that levels of learning are a better predictor of growth than the years-of-schooling measure (Figure 1.2 above). As a result, we would expect that, among a set of simple metrics, the *LAYS* measure should have more explanatory power than the standard years-of-schooling metric does.

We explore this relationship between learning-adjusted school years and growth by

---

[29] We have not attempted to be comprehensive in our discussion here, but instead have stuck to those analyses whose objectives have been most similar to ours. Other related studies include Hanushek and Zhang (2009), who use micro earnings regressions for different cohorts over time to adjust quality by the return to schooling relative to that of a numeraire cohort. Another approach, still further afield from ours, uses micro earnings regressions to do the dual of what we do here—namely, to map increments in learning outcomes back to equivalent years of business-as-usual schooling (Evans and Yuan 2017). Finally, another class of studies aims to adjust measures of learning for the fact that some youth have never started school, or have dropped out before testing (for example, Filmer, Pritchett, and Hasan 2007; Spaull and Taylor 2015).

running a simple cross-country regression using the pooled sample of countries that participated in the 1999 or 2003 rounds of TIMSS (47 countries) or participated in the 2000 round of PISA (42 countries).[30] The results strongly support the notion that *LAYS* captures important aspects of both the quantity and quality of schooling. Average years of schooling of the cohort of 25- to 29-year-olds in 2000 is related to subsequent GDP per capita growth (Column 1 of Table 1), as are average test scores (in mathematics; Column 2). The coefficients on both variables fall substantially when both are included in the model (Column 3), and, unsurprisingly, the explanatory power of the model increases (R-squared of 0.68, compared with 0.65 with average years alone and 0.59 with test scores alone).[31]

LAYS is also strongly related to subsequent growth (Column 4). Importantly, the explanatory power of the model that includes just the *LAYS* measure (R-squared of 0.68) is the same as that of the model that includes years of schooling and test scores as separate variables (R-squared of 0.68), and greater than the model that includes just average years of schooling or just test scores. This means that *LAYS,* as a single variable, captures key dimensions of its constituent parts that are associated with economic growth. The findings are consistent (although point estimates differ because of smaller and different country coverage) when using just the TIMSS or the PISA samples (Columns 5-12): again, *LAYS* explains about as much of the variation as its constituent parts do when entered together. The findings are likewise consistent under various other robustness tests—notably, if we use GDP per capita growth between 1990 and 2016, or between 1995 and 2016 (see Annex Tables A1 and A2); if we use the average years of schooling for the whole population aged 15 and over, together with the corresponding *LAYS* measure (Annex Table A3); and if we use the percent of the student population attaining Level 1 or above on PISA and TIMSS as our measure of learning (Annex Table A4).

Beyond its usefulness for research, another effect of *LAYS* could be to improve the incentives for policymakers—to help keep their focus on what really matters, which is schooling *with* learning. The Sustainable Development Goals reflect a shared international commitment to achieve this. SDG 4 commits countries to ensure education quality and learning for all, and SDG

---

[30] Countries that participate in both (22 countries) appear twice in this pooled sample.
[31] All models include initial GDP per capita, and the pooled models include a dummy variable for whether the test score is from TIMSS or PISA.

4 indicators include measures of learning in primary and lower secondary, among other measures of learning and skills. However, these learning indicators cannot stand alone as policy targets, because they do not capture progress in participation and attainment. A summary macro measure like *LAYS* that captures both could encourage a sustained focus on both quantity and quality of education.

Table 1: Regression of GDP per capita growth 2000-2016 on various measures of education

	TIMSS and PISA				TIMSS only				PISA only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Avg. years of schooling (age 25-29)	0.327		0.204		0.343		0.183		0.288		0.192	
	(5.28)**		(2.84)**		(4.05)**		(1.81)		(3.05)**		(1.77)	
Learning Adjusted Years of Schooling				0.298				0.319				0.264
				(6.27)**				(5.08)**				(3.58)**
Average test score (TIMSS or PISA)		0.973	0.606			1.044	0.685			0.914	0.581	
		(5.29)**	(3.01)**			(4.40)**	(2.56)*			(3.07)**	(1.70)	
Initial GDP per capita (log)	-1.280	-1.248	-1.380	-1.375	-1.255	-1.132	-1.301	-1.316	-1.313	-1.430	-1.483	-1.460
	(11.63)**	(10.75)**	(12.54)**	(12.54)**	(7.92)**	(7.42)**	(8.69)**	(8.99)**	(8.53)**	(7.81)**	(8.23)**	(8.73)**
TIMSS dummy (0/1)	-0.079	-0.331	-0.203	-0.003								
	(0.38)	(1.50)	(1.00)	(0.01)								
Constant	10.783	9.564	10.308	12.539	10.306	7.854	9.231	11.815	11.544	11.632	11.581	13.708
	(12.10)**	(10.26)**	(11.93)**	(14.32)**	(9.01)**	(6.36)**	(8.01)**	(10.73)**	(8.71)**	(9.23)**	(8.95)**	(10.51)**
R^2	0.65	0.60	0.69	0.68	0.61	0.56	0.66	0.66	0.67	0.66	0.70	0.69
N	84	88	84	84	44	47	44	44	40	41	40	40

Data sources: average years of schooling for the cohort of 15 years and over from Barro-Lee. Test scores from 1999 TIMSS and 2000 PISA mathematics assessment. TIMSS from 1999 is augmented with results from 2003 for additional countries. GDP per capita and growth from World Development Indicators.

Notes: * $p<0.05$; ** $p<0.01$. T-statistics in parentheses.

6.2 What LAYS leaves out

No matter how well they are designed, the student assessments from which the internationally comparative learning measures are drawn cannot possibly capture all the learning that takes place in a school, for several reasons. First, the most established international assessments—TIMSS, PISA, PIRLS, and regional assessments like the Latin American Laboratory for Assessment of the Quality of Education (LLECE), the Programme d'Analyse des Systèmes Educatifs de la Confemen (PASEC), and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)—test students in only a handful of disciplines: mathematics, language, and science. This means that *LAYS* does not directly capture learning in other fields, such as history, civics, and the arts. Second, because these assessments are low-stakes sample-based tests, students may not take them seriously, meaning that the test results will underestimate their actual learning. Moreover, if students in some countries take the tests more seriously than their peers in other countries, this will bias cross-country comparisons of learning (Gneezy and others 2017). Third, none of these cognitive-skills measures will directly capture the other classes of skills that students acquire through schooling. These omitted skills include socioemotional (or non-cognitive) skills, such as perseverance, conscientiousness, openness, and emotional stability, and job-specific technical skills, such as programming or welding.

One argument in favor of using *LAYS* is that all these other outcomes of education are already ignored in the standard years-of-schooling metric. *LAYS* at least captures some key learning outcomes that every schooling system is trying to achieve—including literacy, numeracy, and reasoning abilities. It thereby signals that educational progress requires improvements in these learning outcomes. Even if the other outcomes fail to improve, then using *LAYS* could improve incentives for policymakers by inducing them to pay greater attention to quality and learning.

But there remains the possibility that *LAYS* could create perverse incentives, if the increased attention crowds out the other desirable outcomes of schooling. School systems could give less time to unmeasured subjects or socioemotional skills, leaving students less prepared for further schooling, employment, or life in general. Is this a real concern? In many of the

countries with the poorest learning outcomes, it may not be. Steps to improve outcomes on measured cognitive skills may well "crowd in" rather than "crowd out" other skills, for several reasons. First, as the WDR 2018 argues, "[c]onditions that allow children to spend two or three years in school without learning to read a single word, or to reach the end of primary school without learning to do two-digit subtraction, are not conducive to reaching the higher goals of education" (World Bank 2018). Taking steps to improve cognitive learning, for example by ensuring that teachers are in class and well prepared to teach, will likely also improve students' skills and attitudes in other areas, such as conscientiousness, perseverance, and citizenship. Second, learning more in one area may directly improve the student's skills in other cognitive or socioemotional domains. For example, incentives targeting math and language learning have been shown to improve test scores in science and social studies, probably because of positive spillovers from numeracy and literacy that make it easier to learn in general (Muralidharan and Sundararaman 2011). Moreover, "skills beget skills" across the cognitive and socioemotional domains (Cunha and Heckman 2007). For example, improving a child's cognitive competence may increase her self-esteem or citizenship knowledge.

Our point is not that *LAYS* captures everything important. Even if improving cognitive skills will often crowd in other desirable outcomes of schooling, crowding out remains a real possibility. For this reason, countries will need other measures to capture progress in other areas not covered by *LAYS*. But when we need a single simple measure of educational progress for research or policy, *LAYS* may be a very good alternative to the standard measure.

7. Conclusion

In terms of its value for education and human capital, a year of schooling does not mean the same thing in every country. As recent research has emphasized, there are vast discrepancies in what students learn each year across different countries and contexts, especially when we consider not just the high-income countries but also the low- and middle-income countries (Pritchett 2013, World Bank 2018). Yet the standard macro-level measures of educational investment and capital in a society implicitly assume away these differences. By simply counting years of schooling, they equate a year of schooling in Finland or Singapore with a year

of schooling in Mauritania or Mozambique.

In this paper, we have presented a new measure, the Learning-Adjusted Years of Schooling (or *LAYS*), that adjusts quantity for quality transparently in a straightforward macro metric of education. Given the vast cross-country differences in learning, we argue it is useful to have a macro measure like this, for both research and policy purposes. From a research perspective, it is more informative than the standard years-of-schooling measure: for example, we show that *LAYS* explains more of the cross-country variation in GDP growth rates (and as much as years of schooling and test scores when they both appear in the model). And from a policy perspective, it removes the perverse incentive provided by the standard measure, which is to focus on getting children into school without paying attention to learning outcomes. As the experience of the Millennium Development Goals showed, this perverse incentive is not just theoretical.

Beyond its usefulness, *LAYS* is robust to variations in the assessment system and subject that is used to calculate it. This need not have been the case: one of the negative features of *LAYS* is that it is theoretically dependent on the specific learning metric and units used. And yet in practice, calculations drawing on different international assessments, such as PISA and TIMSS, yield very similar results, as do calculations based on verbal, math, and science scores. While this is reassuring, these variations are all based on assessments that use the same units to report results. But as we show, unit-free *LAYS* alternatives (based on the percentage of students in each system who achieve at least a minimum level of learning) also lead to values that are highly correlated with the *LAYS* measures. These alternatives do penalize the schooling of low performers even more than *LAYS* does, so the basic *LAYS* adjustment proposed here should perhaps be seen as a lower bound on the quality adjustments that are necessary. Or, to put it another way, although the *LAYS* measures of quality-adjusted schooling for some countries may seem shockingly low, other plausible alternatives would make schooling in those countries look even worse.

This specific version of *LAYS* is just an illustration. The same approach could be applied to data from other assessments or other measures of years of schooling. But given the prominence of PISA and TIMSS in the global education-policy dialogue, we think this version is

a reasonable starting point, combining as it does major assessments and measures that are already in wide use.[32] Further elaboration on this approach could delve into the *LAYS* scores for specific populations, for example by disaggregating by gender or rural/urban location.

Like all aggregate measures, *LAYS* should be used with caution. We have shown only point estimates, but because there are standard errors around its test measures and years-of-schooling components, any *LAYS* measure will also have some error band around it. This means that it is important not to overinterpret small cross-country differences or small changes over time. But used correctly, a measure like *LAYS* could add considerable value to the policy dialogue. It highlights that cross-country differences in education and human capital are much greater than the standard measure suggests, and it provides an alternative measure for research and policy that captures the reality on the ground much more accurately.

---

[32] For example, see Kraay (2018), which applies this *LAYS* approach in constructing a broader measure of human capital.

References

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*. 106(4): 979–1014.

ASER Centre. 2017. Annual Status of Education Report (Rural) 2016. New Delhi: ASER Centre. http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER%202016/aser_2016.pdf.

Barro, Robert J, and Jong Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950–2010." *Journal of Development Economics* 104: 184-98. Data available at http://www.barrolee.com.

Bedard, Kelly, and Elizabeth Dhuey. 2006. "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects." *The Quarterly Journal of Economics* 121(4): 1437–1472.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129 (4): 1553–1623.

Crawford, Claire, Lorraine Dearden and Costas Meghir. 2010. "When You Are Born Matters: The Impact of Date of Birth on Educational Outcomes in England," DoQSS Working Papers 10-09, Department of Quantitative Social Science - UCL Institute of Education, University College London.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31-47.

Dobkin, Carlos and Fernando Ferreira. 2010. "Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes?" *Economics of Education Review* 29(1): 40-54.

Elder, Todd E. and Darren H. Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers." *Journal of Human Resources*. 44(3): 641-683.

Evans, David and Fei Yuan. 2017. "The Economic Returns to Interventions That Increase Learning." World Development Report 2018 Background Paper, World Bank, Washington, DC.

Fredriksson, Peter, and Bjorn Öckert. 2014. "Life-cycle Effects of Age at School Start." *Economic Journal*. 124: 977-1004.

Filmer, Deon, Amer Hasan, and Lant Pritchett. 2006. "A Millennium Learning Goal: Measuring Real Progress in Education". Center for Global Development Working Paper No. 97.

Gneezy, Uri, John A. List, Jeffrey A. Livingston, Sally Sadoff, Xiangdong Qin, and Yang Xu. 2017. "Measuring Success in Education: The Role of Effort on the Test Itself." NBER Working Paper No. 24004.

Hanushek, Eric A., Jens Ruhose, and Ludger Woessmann. 2017. "Knowledge Capital and Aggregate Income Differences: Development Accounting for US States." *American Economic Journal: Macroeconomics*. 9(4): 184-224.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2015. "Returns to Skills around the World: Evidence from PIAAC." *European Economic Review* 73: 103–30.

Hanushek, Eric A., and Ludger Woessmann. 2012. "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation." *Journal of Economic Growth* 17 (4): 267–321.

Hanushek, Eric A., and Lei Zhang. 2009. "Quality-consistent Estimates of International Schooling and Skill Gradients." *Journal of Human Capital*, *3*(2), 107-143.

Jerrim, John, and Nikki Shure. 2016. "Achievement of 15-Year-Olds in England: PISA 2015 National Report." December. UCL Institute of Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/574925/PISA-2015_England_Report.pdf

Kaarsen, Nicolai. 2014. "Cross-Country Differences in the Quality of Schooling." *Journal of Development Economics*. 107: 215-224.

Kaffenberger, Michelle, and Lant Pritchett. 2017. "The Impact of Education versus the Impact of Schooling: Schooling, Reading Ability, and Financial Behavior in 10 Countries." World Development Report 2018 Background Paper, World Bank, Washington, DC.

Kawaguchi, Daiji. 2011. "Actual Age at School Entry, Educational Outcomes, and Earnings." *Journal of the Japanese and International Economies* 25(2): 64-80.

Kraay, Aart. 2018. "Methodology for a World Bank Human Capital Index." World Bank Policy Research Working Paper No. 8593. The World Bank.

McEwan, Patrick J. and Joseph S. Shapiro. 2008. "The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates." *Journal of Human Resources*. 43(1):1-29.

Mullis, I. V. S., M. O. Martin, P. Foy, and M. Hooper. 2016. "TIMSS 2015 International Results in Mathematics." TIMSS and PIRLS International Study Center, Boston College, Chestnut Hill, MA. http://timssandpirls.bc.edu/timss2015/international-results/.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77.

OECD. 2013. *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed. (Volume II)*. Paris, OECD Publishing. http://dx.doi.org/10.1787/9789264201132-en

OECD. 2016. *PISA 2015 Results (Volume 1). Excellence and Equity in Education*. Paris, OECD. http://dx.doi.org/10.1787/9789264266490-en.'

Oye, Mari, Lant Pritchett, and Justin Sandefur. 2016. "Girls' Schooling Is Good, Girls' Schooling with Learning Is Better." Education Commission, Center for Global Development, Washington, DC.

Pehkonen, J., J. Viinikainen, P. Böckerman, L. Pulkki-Råback, L. Keltikangas-Järvinen and O.

Raitakari. 2015. "Relative age at school entry, school performance and long-term labour market outcomes." *Applied Economics Letters* 22(16):1345-1348.

Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: Center for Global Development.

Psacharopoulos, George, and Harry Anthony Patrinos. 2018. "Returns to Investment in Education: A Decennial Review of the Global Literature." World Bank Policy Research Working Paper No. 8402.

Puhani, Patrick A. and Andrea Maria Weber. 2007. "Persistence of the School Entry Age Effect in a System of Flexible Tracking" IZA Discussion Paper No. 2965; U of St. Gallen Economics Discussion Paper No. 2007-30. Available at SSRN: https://ssrn.com/abstract=1005575

Robertson, Erin. 2011. "The Effects of Quarter of Birth on Academic Outcomes at the Elementary School Level." *Economics of Education Review* 30(2):300-311.

Schoellman, Todd. 2012. "Education Quality and Development Accounting." *The Review of Economic Studies*. 79(1): 388–417. https://doi.org/10.1093/restud/rdr025

Singh, Abhijeet. 2017. "Learning More with Every Year: School Year Productivity and International Learning Divergence." Manuscript.

Spaull, Nicholas, and Stephen Taylor. 2015. "Access to What? Creating a Composite Measure of Educational Quantity and Educational Quality for 11 African Countries." *Comparative Education Review* 59:1: 133-165.

Valerio, Alexandria, Maria Laura Sanchez Puerta, Namrata Tognatta, and Sebastian Monroy-Taborda. 2016. "Are There Skills Payoff in Low- and Middle-Income Countries? Empirical Evidence Using STEP Data." World Bank Policy Research Working Paper No. 7879. World Bank, Washington, DC

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC.

Annex 1: Additional results

Annex 1 Figure 1: Implications of using share reaching TIMSS intermediate benchmark for LAYS adjustment
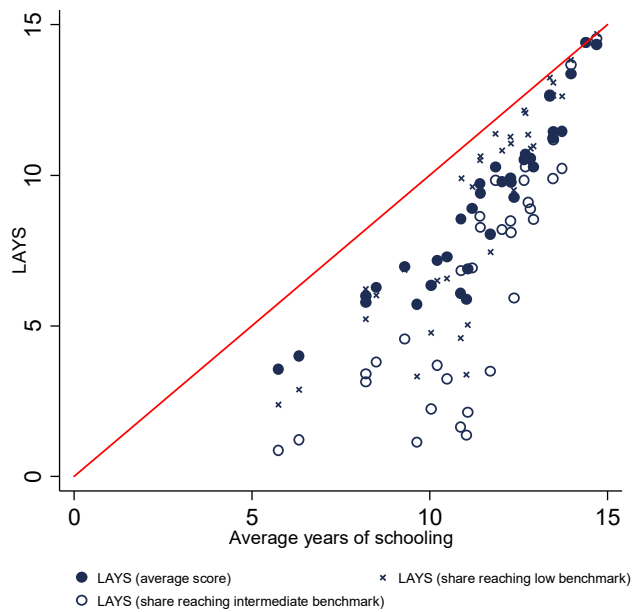
Source: Authors' analysis of TIMSS 2015 and Barro-Lee data.
Notes: Numeraire country is Singapore. Illustration for case where learning starts at Grade 0.

Annex 1 Figure 2: Implications of using share reaching PISA Level 1 or Level 2 benchmark for LAYS adjustment
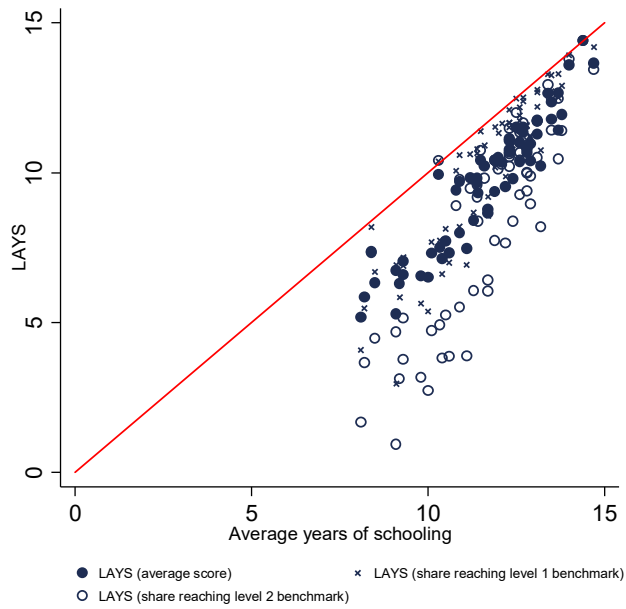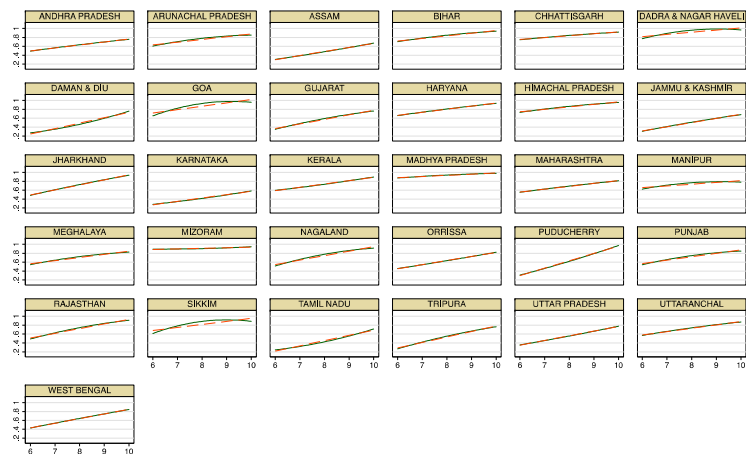

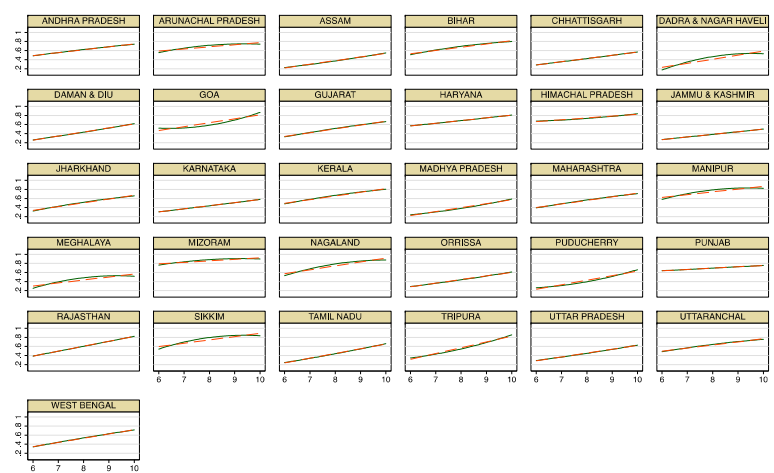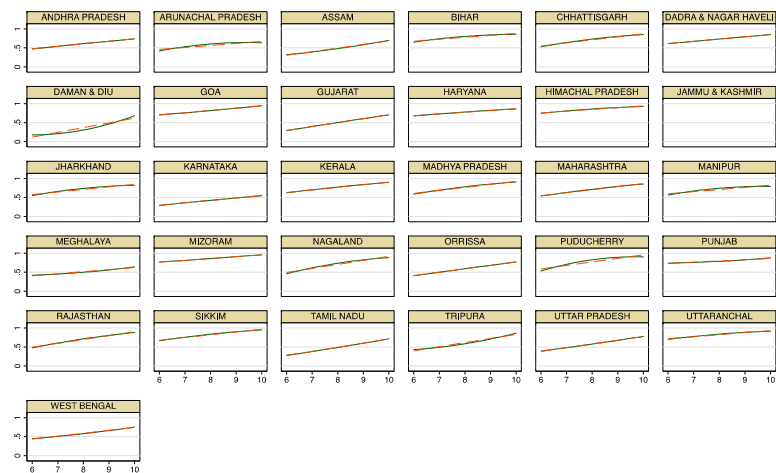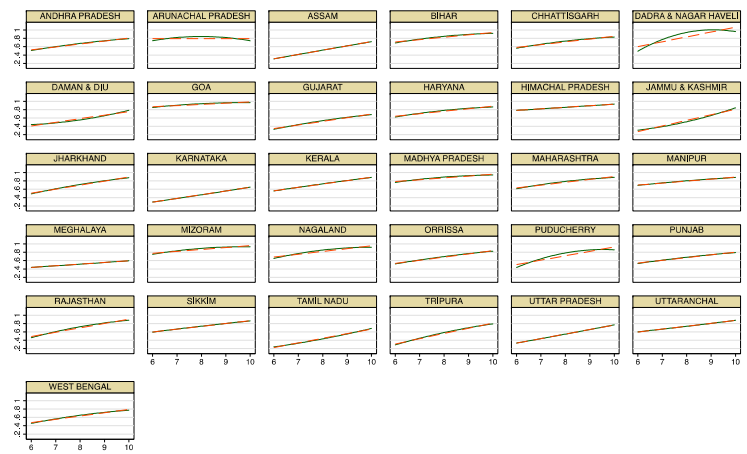Source: Authors' analysis of PISA 2015 and Barro-Lee data.
Notes: Numeraire country is Singapore. Illustration for case where learning starts at Age 6.

Annex 1 Figure 3: Learning Trajectories, Indian States. Proportion who master b*asic math skills, Grades 6-10, Linear/Curved Fit*

2008	2009

Annex Table A1: Regression of GDP per capita growth 1995-2016 on various measures of education

	TIMSS and PISA				TIMSS only				PISA only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Avg. years of schooling (age 25-29)	0.322		0.222		0.335		0.204		0.295		0.222	
	(5.81)**		(3.38)**		(4.47)**		(2.22)*		(3.41)**		(2.18)*	
Learning Adjusted Years of Schooling				0.280				0.301				0.247
				(6.48)**				(5.33)**				(3.58)**
Average test score (TIMSS or PISA)		0.874	0.483			0.932	0.556			0.817	0.422	
		(5.32)**	(2.59)*			(4.51)**	(2.27)*			(2.88)**	(1.30)	
Initial GDP per capita (log)	-1.012	-0.977	-1.091	-1.087	-1.006	-0.897	-1.042	-1.048	-1.021	-1.106	-1.145	-1.142
	(10.06)**	(9.32)**	(10.72)**	(10.69)**	(7.06)**	(6.71)**	(7.62)**	(7.87)**	(7.07)**	(6.23)**	(6.66)**	(7.12)**
TIMSS dummy (0/1)	-0.085	-0.299	-0.186	-0.014								
	(0.45)	(1.50)	(0.99)	(0.08)								
Constant	8.244	7.431	7.863	9.912	7.968	6.143	7.089	9.379	8.640	8.942	8.680	10.764
	(10.08)**	(8.83)**	(9.79)**	(12.25)**	(7.69)**	(5.66)**	(6.68)**	(9.37)**	(6.93)**	(7.34)**	(7.03)**	(8.64)**
R^2	0.57	0.52	0.61	0.60	0.55	0.51	0.60	0.60	0.58	0.54	0.60	0.59
N	83	87	83	83	44	47	44	44	39	40	39	39

Data sources: average years of schooling for the cohort of 15 years and over from Barro-Lee. Test scores from 1999 TIMSS and 2000 PISA mathematics assessment. TIMSS from 1999 is augmented with results from 2003 for additional countries. GDP per capita and growth from World Development Indicators.

Notes: * $p<0.05$; ** $p<0.01$. T-statistics in parentheses.

Annex Table A2: Regression of GDP per capita growth 1990-2016 on various measures of education

	TIMSS and PISA				TIMSS only				PISA only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Avg. years of schooling (age 25-29)	0.140		0.042		0.119		-0.026		0.148		0.079	
	(2.32)*		(0.63)		(1.47)		(0.30)		(1.57)		(0.74)	
Learning Adjusted Years of Schooling				0.162				0.165				0.154
				(3.45)**				(2.77)**				(2.04)*
Average test score (TIMSS or PISA)		0.592	0.531			0.611	0.661			0.609	0.445	
		(3.70)**	(2.89)**			(3.34)**	(3.02)**			(2.02)	(1.31)	
Initial GDP per capita (log)	-0.674	-0.721	-0.780	-0.773	-0.576	-0.627	-0.633	-0.690	-0.775	-0.865	-0.915	-0.881
	(5.98)**	(6.81)**	(6.89)**	(6.83)**	(3.58)**	(5.05)**	(4.35)**	(4.70)**	(4.84)**	(4.49)**	(4.79)**	(4.97)**
TIMSS dummy (0/1)	-0.145	-0.234	-0.229	-0.092								
	(0.74)	(1.24)	(1.21)	(0.49)								
Constant	6.821	6.045	6.453	7.802	6.014	4.869	5.015	6.925	7.701	7.351	7.746	8.931
	(8.12)**	(7.57)**	(7.98)**	(9.07)**	(5.65)**	(5.13)**	(4.97)**	(6.51)**	(5.93)**	(5.83)**	(6.03)**	(6.69)**
R^2	0.37	0.40	0.44	0.42	0.31	0.42	0.46	0.41	0.43	0.41	0.46	0.45
N	73	76	73	73	36	38	36	36	37	38	37	37

Data sources: average years of schooling for the cohort of 15 years and over from Barro-Lee. Test scores from 1999 TIMSS and 2000 PISA mathematics assessment. TIMSS from 1999 is augmented with results from 2003 for additional countries. GDP per capita and growth from World Development Indicators.

Notes: * $p<0.05$; ** $p<0.01$. T-statistics in parentheses.

Annex Table A3: Regression of GDP per capita growth 2000-2016 on various measures of education

	TIMSS and PISA				TIMSS only				PISA only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Avg. years of schooling (age 15+)	0.309		0.196		0.305		0.148		0.305		0.229	
	(4.80)**		(2.92)**		(3.32)**		(1.54)		(3.31)**		(2.37)*	
Learning Adjusted Years of Schooling				0.324				0.337				0.306
				(6.04)**				(4.58)**				(3.89)**
Average test score (TIMSS or PISA)		0.973	0.686			1.044	0.791			0.914	0.610	
		(5.29)**	(3.71)**			(4.40)**	(3.21)**			(3.07)**	(1.99)	
Initial GDP per capita (log)	-1.228	-1.248	-1.381	-1.357	-1.173	-1.132	-1.278	-1.283	-1.301	-1.430	-1.508	-1.471
	(11.23)**	(10.75)**	(12.60)**	(12.31)**	(7.26)**	(7.42)**	(8.55)**	(8.47)**	(8.85)**	(7.81)**	(8.59)**	(9.11)**
TIMSS dummy (0/1)	-0.075	-0.331	-0.220	-0.011								
	(0.35)	(1.50)	(1.09)	(0.05)								
Constant	10.960	9.564	10.337	12.558	10.423	7.854	9.120	11.770	11.722	11.632	11.630	13.835
	(12.04)**	(10.26)**	(11.99)**	(14.13)**	(8.67)**	(6.36)**	(7.87)**	(10.26)**	(9.08)**	(9.23)**	(9.36)**	(10.81)**
R^2	0.63	0.60	0.69	0.68	0.57	0.56	0.66	0.64	0.68	0.66	0.71	0.71
N	84	88	84	84	44	47	44	44	40	41	40	40

Data sources: average years of schooling for the cohort of 15 years and over from Barro-Lee. Test scores from 1999 TIMSS and 2000 PISA mathematics assessment. TIMSS from 1999 is augmented with results from 2003 for additional countries. GDP per capita and growth from World Development Indicators.

Notes: * $p<0.05$; ** $p<0.01$. T-statistics in parentheses.

Annex Table A4: Regression of GDP per capita growth 2000-2016 on various measures of education

	TIMSS and PISA				TIMSS only				PISA only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Avg. years of schooling (age 25-29)	0.327		0.175		0.343		0.199		0.288		0.026	
	(5.28)**		(2.26)*		(4.05)**		(1.95)		(3.05)**		(0.18)	
Learning Adjusted Years of Schooling				0.235				0.226				0.262
				(6.47)**				(5.08)**				(3.79)**
Pct. lvl 1 and above (TIMSS or PISA)		3.294	2.155			2.971	1.857			5.451	5.143	
		(5.38)**	(3.03)**			(4.12)**	(2.31)*			(3.78)**	(2.30)*	
Initial GDP per capita (log)	-1.280	-1.163	-1.298	-1.307	-1.255	-1.094	-1.285	-1.282	-1.313	-1.438	-1.437	-1.369
	(11.63)**	(9.80)**	(11.15)**	(11.72)**	(7.92)**	(7.16)**	(8.50)**	(9.01)**	(8.53)**	(7.14)**	(7.02)**	(7.20)**
TIMSS dummy (0/1)	-0.079	0.190	0.136	0.164								
	(0.38)	(0.78)	(0.60)	(0.77)								
Constant	10.783	10.387	10.762	12.301	10.306	10.190	10.745	12.312	11.544	11.245	11.206	12.644
	(12.10)**	(10.06)**	(11.29)**	(12.75)**	(9.01)**	(8.36)**	(9.73)**	(10.88)**	(8.71)**	(7.13)**	(6.93)**	(7.80)**
R^2	0.65	0.59	0.67	0.68	0.61	0.54	0.66	0.66	0.67	0.63	0.63	0.63
N	84	80	77	77	44	47	44	44	40	33	33	33

Data sources: average years of schooling for the cohort of 25- to 29-year-olds from Barro and Lee (2013). Test scores from 1999 TIMSS and 2000 PISA mathematics assessment. TIMSS from 1999 is augmented with results from 2003 for additional countries. GDP per capita and growth from World Development Indicators.

Notes: * $p<0.05$; ** $p<0.01$. T-statistics in parentheses.