Learning Outcomes Analysis and Estimates

Annababette Wils

Supporting Paper for the UNFPA Project (RFP ID UNFPA/USA/14/037) Investment in Improving Health, Education and other Outcomes for Adolescents

Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne
March 2019

Learning is the purpose of education, and perhaps naively, many thought, up until recently, that if children attended school, they would be learning. Unfortunately, mounting evidence – emerging particularly in the past 10 years – shows learning levels in developing countries that are far below what they should be, with many students not able to read simple sentences or do basic addition and subtraction even after years of schooling[1].

Even as this general insight has emerged, it has also become clear that there are many gaps in this knowledge of how well children are learning in low- and middle-income countries. There are, to date, many countries for which no public assessments of learning exist, and, among those that exist, there is not an agreed-upon methodology to make them comparable.

This paper provides efforts to compile available learning data, to standardize the outcomes, and to make estimates of learning outcomes for countries where no international learning assessments exist. In the future, we believe more countries will have standardized assessments, but in the interim, estimates can be useful.[2]

Learning is at the heart of education. Estimates suggest that learning levels in developing countries are far below what they should be.[3] Learning is a complex process and there is not a precise agreement on exactly what children should learn or on how it should be measured. There is general agreement that, at least at the primary level, it should include reading, writing, and mathematics skills.

Unlike the generally agreed-upon definitions of completion (with the exception of some gray areas for students who attend only part of a school year) there is not an agreed-upon definition of learning. Learning is measured annually in many, if not most, countries for specific grades through entrance, certification, and benchmark exams. These assessments have not been made internationally comparable and sometimes are not even comparable from one year to the next within a particular country.[4]

There are a growing number of *sample-based assessments,* many of which are applied across countries in a way that makes it possible to compare the outcomes internationally. These assessments include global series such as the TIMSS, PIRLS, and PISA tests; and regional series such

---

[1] See, for example, the most recent GEMR report Chapter 10, UNESCO 2016, Education for People and Planet: Creating Sustainable Futures For All, Global Education Monitoring Report, 2016, Paris.
[2] The UNESCO Institute for Statistics (UIS) is currently leading in the development of these methodologies.
[3] See, for example, the most recent GEMR report Chapter 10, UNESCO 2016, cited above.
[4] See, for example, EPDC 2008 for an analysis of national learning assessments, Education Policy and Data Center 2008, Sub-national Disparities in Learning: An Analysis of Differentials in Learning Scores in 25 Countries, and the Correlations with School Entry and Retention Rates and Pupil-Teacher Ratios, EPDC Working Paper No. EPDC-08-10, FHI360, Washington DC.

as the LLECE, SACMEQ, PASEC, ASER, Uwezo, and EGRA/EGMA tests.[5] Taken together, these assessments covered about 55 low- and middle-income countries at the primary school level (the exact number depends on which assessments are included); and 29 middle-income countries at the secondary school level between 2010 and 2015.[6]

Learning is an extremely complex and rich process and does not lead to a binary outcome like completing primary school – or not. Knowing this, we still seek to distill learning into one, or a few, simple measures so that it's possible to compare a larger number of countries or to analyze trends. The two measures typically used for this are average assessment scores or the percentage of children reaching a set minimum benchmark.

Many analysts in the past have used *average scores* to compare countries.[7] Average scores, one can argue, capture the *overall performance* of a system. But from an equity perspective, average scores are less useful. The same average score can represent a highly unequal system with one pool of students not learning at all and a second pool of high achievers, or it can represent a system where everyone is reaching at least a minimum level of learning. The Commission's commitment to equity and to a basic quality of education *for all* led us to conclude that minimum benchmarks are better markers for our purposes.

The UNESCO GEMR has been using the ***percentage of children who reach minimum levels of learning in primary school and in lower secondary*** as the cut-off measures for learning in its online WIDE database.[8] This is a sensible approach because it signals that there is a benchmark of skills that all children should acquire, and particularly at the primary level, mastering reading is useful because children need to be able to read in order to achieve further learning. This is the measure that we will discuss in detail in the remainder of this section.

1. Review of existing international datasets of learning that combine assessments

In order to maximize the number of countries for which we have indicators on the achievement of basic benchmarks, we need to combine the different assessment series. Each assessment has its own metrics, and so to standardize the scores, they need to be anchored to each other. There is not yet a standardized method for anchoring the assessments through variations have been proposed. The basic approach is the following: One assessment series is chosen as the anchor. Then analysts use countries that took the anchor assessment as well as one of the other assessments to develop a conversion factor. This allows all of the results from the other assessment(s) to be expressed in the units of the anchor assessment. There are two predominate general methods for this anchoring exercise – the first is to use average scores and the second to match items in the assessments. Existing work in this area is outlined below.

---

[5] Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), and the Programme for International Student Assessment (PISA) tests; and regional series such as the Latin American Laboratory for Assessment of the Quality of Education (LLECE), the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), Analysis Programme of the Confemen Education Systems (PASEC), the Annual Status of Education (ASER), Uwezo, and EGRA/EGMA tests.

[6] School grades per level vary by country, but for consistency, we take grade 6 as the last year of primary school. The SACMEQ test is given to grade 6 students.

[7] See, for example, Hanushek, EA and DD Kimko 2000, 'Schooling, labor force quality and the growth of nations', American Economic Review, 90(5): 1184-1208; Altinok, N and Murseli, H 2007, 'International database on human capital quality', Economics Letters, 96(2): 237-244); Beatty, A and Pritchett, L 2012, From Schooling Goals TO Learning Goals: How Fast Can Student Learning Improve?, CGD Policy Paper 012, Washington DC: Center for Global Development, http://www.cgdev.org/content/publications/detail/1426531, updated Jan 29, 2013.

[8] UNESCO World Inequalities Database on Education, Paris, http://www.education-inequalities.org/

The basic methodology used to create the GEMR dataset dates back to the early 2000s. Hanushek and Kimko created an early international dataset.[9] Altinok and Murseli built on their approach and proposed a revised method and created a new dataset.[10] Later, Altinok[11] and Angrist et al.[12] built on the Altinok and Murseli method to create two international databases of learning scores – UNESCO's WIDE database noted earlier, and the World Bank Global Dataset on Education Quality. The former includes percentages of students who reach at least reading benchmarks, and the latter provides standardized average scores.

Up until now the basic methodology for standardizing minimum learning outcomes, used by UNESCO and the UNESCO Institute for Statistics, and the World Bank is based on the methodology developed by Altinok (2012).[13] Altinok's method uses average scores to anchor and standardize the outcomes, which are then converted to standardized percentages of children reaching the minimum benchmarks. Standardizing *average* scores is difficult and we developed a simpler, and potentially more accurate methodology as described below.

Methodology for creating a dataset of standardized learning scores for countries with international learning assessments

This section describes our methodology for standardizing learning outcomes based on a selected minimum benchmark level and multiple international learning assessments series. Two steps are described: first, the selection of the minimum benchmark, and second, aligning multiple assessments.

Selection of the benchmarks for minimum learning

A useful starting point for this is the group of high-performing countries, most of them industrialized. Education can and will improve even among these countries, but at present, this group represents the best learning outcomes in the world and, if we believe in the convergence of education for all, these levels should be the minimum aspiration for developing countries.

The TIMSS, PIRLS and PISA assessment results from these countries suggests that for each assessment, the "low" level of learning is a useful baseline benchmark that all – or nearly all – students should reach. Even in the high-performing countries, some percentages of students miss these benchmarks. For example, between 1 and 7 percent of students in high-income countries participating in the PIRLS assessment did not reach those levels for reading. This indicates that the measure is meaningful and is not a non-measure of absolutely no learning at all -- if it were, 100 percent of pupils would attain at least this level. The "intermediate" levels are missed by far larger portions of pupils even in high-income countries, so using these levels as the aspirational target for all pupils including those in low- and middle-income countries would set the bar beyond what even the best systems today are able to achieve today.

---

[9] Hanushek and Kimko 2000, cited above.
[10] Altinok and Murseli 2007, cited above.
[11] Altinok, N 2012, A New International Database on the Distribution of Student Achievement, Background paper prepared for the Education for All Global Monitoring Report 2012 Youth and Skills: Putting Education to Work*,* UNESCO, Paris.
[12] Hanushek and Kimko 2000, cited above; Altinok and Murseli 2007, cited above; Altinok 2012, cited above; UNESCO 2012, Youth and Skills: Putting Education to Work, EFA Global Monitoring Report, Paris; Angrist, N, Patrinos, HA and Schlotter, M 2013, An Expansion of Global Data Set on Educational Quality: A Focus on Achievement in Developing Countries, Policy Research Working Paper Series 6536, Washington DC: World Bank, http://datatopics.worldbank.org/Education/files/GlobalAchievement/Full.pdf.
[13] Altinok. 2012, cited above.

The assessments include different subjects – reading, math, and science. Of these, we selected reading benchmarks reaching at least the PIRLS "low" level or equivalent as the primary school learning measure; and a math score reaching at least the PISA "level 1" or equivalent for the secondary school measure.

At the primary level, we selected the reading score because we believe that literacy is of such basic importance to survive and thrive in today's world, but also because the ability to read is the foundation of all further learning in school. The math score for secondary was selected for two practical reasons: first, the desire to include the TIMSS grade 8 assessment, which measures only math and science; and second, the idea that the mathematics curricula are more standardized, i.e., cover addition, subtraction, fractions, algebra, and so forth, than science curricula and therefore mathematics scores are less likely to be influenced by the selection of topics covered in the assessments.

Aligning multiple assessment series.

As mentioned, currently, the methodology for standardizing the learning outcomes used by UNESCO and UNESCO Institute of Statistics is based on Altinok.[14] This method uses average scores to anchor and standardize the outcomes, and then converts the average scores to standardized percentages of children reaching the minimum benchmarks.

Standardizing average scores is not straightforward. For example, say we have assessment A taken by one set of countries and assessment B taken by a different group, with one country that took both. Say that in this overlapping country, the average score was 300 on assessment A, and 500 on assessment B. Both were taken by a representative set of pupils, so we can infer that an average score of 300 on assessment A is roughly equivalent to an average score of 500 on assessment B, a ratio of 5:3. Unfortunately, it is not possible to simply take that ratio and apply it to all the other average scores of assessment A to create equivalents of assessment B because we can't assume that each additional point in A is worth a 5:3 ratio of a point in assessment B. An accurate transformation requires standardizing the value of *each point* in the anchor and the value of each point in the other assessment so that any average score across the entire *range* can be matched. To do this, we need observations from at least two years from each of the assessments from the anchor country or countries.

On the other hand, if the measure of interest is *percentage of children who reach a particular benchmark*, matching can be much simplified. Instead of a *range* that needs to be matched, we can use just *one point of matching* – the benchmark level of learning. Only one observed year from each of the two assessments to be matched and one country are necessary, although multiple matching countries are better.

The thinking behind our method is basically: if N percent of students in assessment A reach the learning level A1 in subject S, and an equal N percent of students in assessment B reach the learning level B1 in subject S, then the level of learning measured by A1 is roughly equal to the level measured by B1. All we need are the percentages of students (N) who reach the learning levels A1, B1 etcetera. This of course, requires some leap of faith, but essentially the same assumption is made in the existing methodologies.

[14] Altinok 2012, cited above.

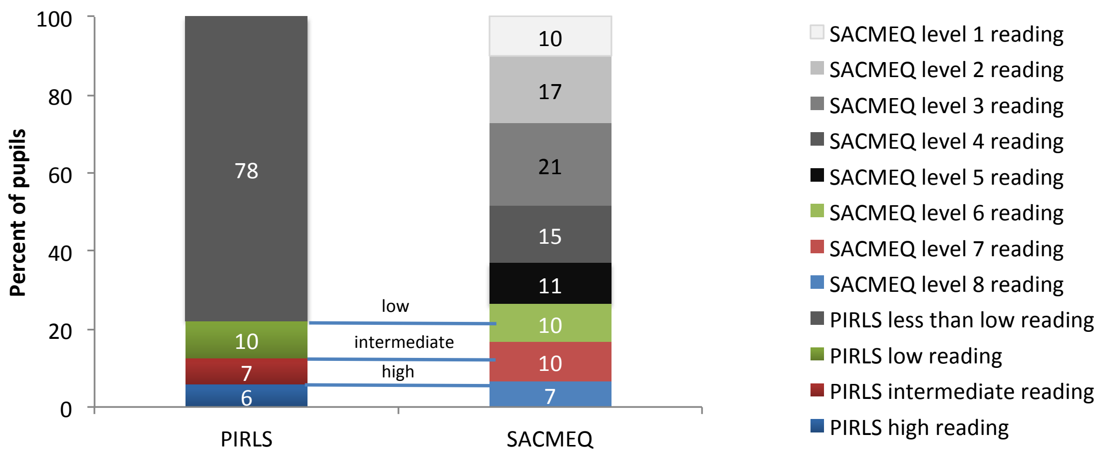
The benchmarks can be measured as a particular cut-off score. For example, for PISA, the cut-off level for reaching a “low” level on the mathematics portion is a score of 420. In other assessments, e.g. EGRA and UWEZO, the benchmark is a particular skill, such as the ability to reach 40 word per minute. Both of these are simplifications of what a child actually knows. For example when we consider a reading assessment, one child may be faster at reading words, while another is faster at grasping the meaning of the words; but it is not really meaningful or efficient to rank one skill above the other. That said, a group of children taking an assessment with a numerical outcome, will end up with set of percentages of children reaching particular scores or skills, and we assume we can use these numerical outcomes as a proxy for what a child has learned.

Another important assumption is that rankings for individuals within a group taking a subject test will remain relatively consistent across assessments. For example, if a group of children take two reading assessments, we assume that each child will end up in roughly the same place in the ranking for the two assessment outcomes compared to other children in the group. Again, this is a simplification, as one reading assessment may focus more on interpretation while another focuses on reading speed, and children may be distributed differently across various skills. Despite these potential imperfections, this assumption is not unreasonable and is necessary for matching exercises.

With these assumptions, we can compare the two sets of rankings from the assessments and match the *percentages of children* who reach each outcome level. **Figure 1** gives an example using outcomes from the PIRLS assessment in 2006 and the SACMEQ assessment of 2007 in South Africa. The figure shows two colored bars, each with the percentage of pupils who reached the reading levels presented by the PIRLS and the SACMEQ assessments respectively; both groups of students were selected to be nationally representative.

All of the assessments we use – TIMSS, PISA, PIRLS and so forth -- , publish tables showing the percentage of students who reach different sets of scores or benchmarks. For example, PIRLS results are presented as the percentage of student who reach the low, intermediate, and high levels of reading. SACMEQ on the other hand presents results for eight levels of reading ability: 1 - Pre-Reading; 2 - Emergent Reading; 3 - Basic Reading; 4 - Reading for Meaning; 5 - Interpretive Reading; 6 - Inferential Reading; 7 - Analytical Reading; and 8 - Critical Reading.

Figure 1. Comparison of the reading level outcomes in South Africa based on PIRLS 2006 and SACMEQ 2007 assessment rankings.



Source: EdStats extraction for learning levels. Figure located in “Learning TIMSSPIRLS Benchmarks 1995-2011.xls” tab “Matching”.

The figure shows that 6 percent of the pupils who took the test reached the PIRLS high level of reading, and 7 percent of pupils who took the SACMEQ assessments reached the level 8. Since the percentages of students reaching PIRLS "high" and SACMEQ 8 are so similar *we infer that the SACMEQ level 8 reading is roughly the equivalent of the PIRLS "high" level*. Thirteen percent of pupils reached at least the PIRLS "intermediate" level of reading; while 17 percent reached at least SACMEQ level 7; 23 percent of pupils reached at least the PIRLS low level of reading while 27 percent reached at least SACMEQ level 6. From this we infer further, that the PIRLS "intermediate" level is roughly equivalent to SACMEQ level 7; and finally, that the PIRLS "low" level is roughly equivalent to SACMEQ level 6 (It's useful to note that at the lower bound of reading skills, the PIRLS assessment does not differentiate, while the SACMEQ assessment has five lower reading levels allowing for more granular categorization of lower reading levels).

Taking SACMEQ reading level 6 as approximately equivalent to the PIRLS low reading level, it is possible transform the learning outcomes of the other 13 Sub-Saharan countries that participated in SACMEQ in 2007, by setting the percentage of pupils who reach SACMEQ reading level 6 as the approximation of the percent who would reach at least the "low" PIRLS reading benchmark. The SACMEQ scores from 2000 were transformed in the same way and reading level 6 was assumed to approximate a PIRLS "low" reading score.

Methodology for creating a dataset of standardized learning scores for countries with no learning assessments

Even with the above alignment of multiple assessments, many countries remain for which there are no observations. For countries with no scores, we developed a predictive model based on other variables that correlate with secondary learning outcomes. The model assumes only not causality correlation – as in, "higher levels of champagne consumption correlate with weddings". We present the methodology and the outcomes for secondary learning levels here. A similar approach can be used to estimate primary learning outcomes.

At the secondary level there are two international assessments -- TIMSS grade 8 (TIMSS8) and the PISA assessment of 15 year-olds. Both the PISA and TIMSS are highly respected global assessment series. Among the countries that participated in both PISA and TIMSS8, learning outcomes are highly aligned; there is a correlation of .91 among the 27 countries that took both of the most recent assessments, so the alignment methodology described above, will cause little shift in the scores regardless of which assessment is used as the standard. Both assessments have respected reputations and either can be reasonably selected as an anchor. We selected the "low" PISA learning outcome as the standard in this paper.

For the middle-income countries that had participated in a PISA or TIMSS assessment in 2007 or later (most observations are from the 2015 round of assessments), it was simplest to take the percentage of students reaching the "low" learning benchmarks in either assessment. The scores are highly correlated in doubloon countries (those that participated in both assessments) with a correlation of 0.93, although on average, slightly fewer pupils reach the PISA "low" level in mathematics, compared to pupils in the same countries who reach the "low" mathematics level in TIMSS. Using PISA as the standard score means that the countries that participated in TIMSS may be given slightly higher scores than they would have in a PISA assessment. The PISA and TIMSS scores provide an assessment of learning for 27 middle-income countries, leaving many middle-income countries and all low-income countries without any standard, international estimate of adolescent learning.

A total of 78 countries participated in either the TIMSS 2011 assessment of the and PISA 2012 round. Of these, 48 were high-income countries, 20 upper-middle income, 9 lower-middle income and zero low-income. Our concern is with the many countries that did not participate.

Earlier, related work on learning focused on identifying the *determinants* of learning outcomes, particularly at the sub-national scale, part of which has been used to create learning production functions. Some of this literature was summarized in the UNESCO 2008 Global Monitoring Report on learning.[15] Hanushek and Krueger analyzed school-level inputs as determinants of learning outcomes.[16] In 2011, the Education Policy and Data Center (EPDC) proposed a production function to predict learning outcomes using a multi-stage regression analysis.[17]

Creating a model of learning outcomes based on contextual variables

Our concern is not necessarily with determinants; for our modeling, variables that correlate with learning outcomes suffice. The model is built from regressions on countries with observed learning outcomes. We then use the coefficients from our model to predict learning outcomes for countries without actual observations, noting that the estimation includes low-income countries with zero actual observations.

We tested a set of socio-economic variables suggested by the earlier work and selected only the small sub-set that significantly improved the model. In addition to the variables from earlier work, we added some of our own, as described below. Some of the selected variables may determine learning outcomes, while others are simply markers of a socio-economic context that correlates with earning outcomes. Data on the socio-economic context variables were compiled from the World Development Indicators and the Barro-Lee dataset on adult education attainment. Table 1 shows the results of our models.

The most significant socio-economic markers we found are: log of GDP per capita; indicators of adult education attainment measured by average years of schooling and percent with secondary education; oil and extraction rent as a percent of GDP. In addition, the learning scores at the primary level were significantly correlated with learning scores at the secondary level. The inclusion of *both* education variables – adult education and primary learning scores does not improve the model fit; but, making two separate models – one with adult education, and one with primary learning scores – increases the number of countries for which there are predictions (because there are some countries for which only one or the other variable is available).

GDP per capita and education – attainment or early learning – are not surprisingly correlated with learning outcomes, as they typically are correlated to higher adult human capital. But, in some countries, a large proportion of GDP (and GDP per capita) is derived from the extraction of natural resources, unrelated to human capital. Oil producing countries were typically found to have much

[15] UNESCO 2008, Education for All by 2015: Will We Make It? EFA Global Monitoring Report, 2008, Paris.
[16] Hanushek, EA 1995, 'Interpreting recent research on schooling in developing countries', World Bank Research Observer, 10(2): 227-246; Hanushek, EA 1997, 'Assessing the effects of school resources on student performance: an update', Educational Evaluation and Policy Analysis, 19(2): 141-164; Hanushek, EA and Luque, JA 2003, 'Efficiency and equity in schools around the world', Economics of Education Review, 22: 481-502; Krueger, A 2003, 'Economic considerations and class size', Economic Journal, 113(485): F34-F63.
[17] Education Policy and Data Center 2011, Estimating Education Output in Developing Countries, EPDC Working Paper 2011-1, FHI-360, Washington DC.

lower learning scores than predicted by their GDP and adult education alone. By including extraction rent as a percent of GDP, we have include a correction factor. Together, these four variables explain 61 percent of the variation in observed secondary learning scores. Other development variables were considered, for example, education survival rates to grade 5 and pupil teacher ratios, but these had almost no effect on improving the model.

After looking at the country-specific model predictions based on these four variables, we observed that the learning outcomes were significantly under-estimated in the set of low-GDP countries that has a socialist or communist past or present. Without commenting on the merits of demerits of these systems, our simple observation is that learning outcomes in these countries are significantly higher than predicted by the GDP and adult learning. When present or former status as a socialist state is included as a dummy variable, the percentage of variation explained by the model increases from an R square of .61 to an R square of .71. Finally, even after these corrections, one country, Vietnam remains as an exceptional outlier. By including a dummy for Vietnam (essentially running the model without Vietnam for all non-Vietnam countries), the explanatory R square rises to .78.

Table 1 also shows an alternate version of each model including only those countries which are in all three models; the coefficients within model appear to be relatively robust to this additional filter.

Each of the variables is correlated with a double-digit range of effects on the percentage of pupils reaching the learning outcomes. In model 4, each tripling of GDP per capita is correlated with, on average, a 10-point increase in the percent of pupils reaching the learning benchmarks.[18] Each 10-percentage point increase in GDP from extraction is correlated with a 7-point decline in learning; in countries with a high portion of income from resources extraction, this this effect can essentially offset the predicted effects of higher incomes. In some countries, extraction makes up as much as 30 to 50 percent of GDP so the offset of learning outcomes can be quite large. On average, each year of adult schooling correlates with 2 additional learning points. Years of adult schooling in the dataset ranges from an average of 2 years in Niger to 14 years in some industrialized countries, meaning differences in adult education can correlate to as much as one quarter of children reaching the learning benchmarks.[19] The often-observed positive effect of socialism on learning outcomes is evident here in the 12-point effect of being a present or former socialist state. Finally, even with all of these adjustments, the PISA results from Vietnam fall far outside the model prediction, making Vietnam an outlier. Including a dummy for Vietnam improves the "fit" of the model from an R-square of .71 to .78, which is a significant improvement. Additionally, "being Vietnam" is correlated with an additional 21 percent of pupils reaching the mathematics learning benchmarks.

[18] (The difference ln(3X) - ln(X) 1).

[19] (Range difference is 12 years, 2 percentage point increase for each year, equals 24 points).

Table 1

Results of selected multivariable models to predict country-level secondary learning levels based on correlation with other country-level indicators, multivariable OLS regressions run with Stata.

Specifications[20]	(Model 1)	(Model 2)	(Model 3)	(Model 1B+)	(Model 2B+)	(Model 3B+)
1 Ln GDP per capita	8.7**	12.4***	17.0***	7.7**	9.4***	15.9***
2 Extraction rent as a percent of GDP	-.57***	-.78***	-.93***	-.57***	.77***	-.97***
3 Percent reaching primary learning benchmark+	.49***			.49***		
4 Average years of schooling adults 25–34		3.2***			3.6***	
5 Socialist/communist state dummy	3.4	11.8***	12.6***	5.0*	9.0**	11.9***
6 Vietnam dummy	26.1 **	40.8***	38.1***	22.8*	38.0***	35.4**
Constant	-46.7*	-75.9***	-88.5***	-35.5*	-48.3*	-74.8**
N	54	69	75	51	51	51
R squared	.78	.78	.71	.81	.74	.66
Root MSE	7.95	8.34	9.33	7.74	8.74	9.83

Notes: *** significant <.1%; ** at <1%; * at <5% level.

+ Only countries that are also included in both models 1 and 2.

Source: Authors' analysis.

Creating secondary learning estimates based on regressions.

The final step is to create a predicted learning outcome for those countries where direct measures are not available. In a general form, the model is: $L = c + \sum_f \beta_f * F_f$, where F denotes the factors of learning, β_f the regression coefficient, and c the constant.

It is possible to make learning outcome predictions based on any of the three models (1), (2), and (3). However, it might be more desirable to create a set of learning estimates that selects the *best possible* model for each country. This would mean, for example, that for countries with GDP per capita, rent extraction, and primary learning outcomes, we use model (1); for countries with GDP per capita, rent extraction, and adult education levels available, we use model (2); and for a third set of countries with neither of the two education observations, we use model (3).

Figure 2 shows observed secondary learning outcomes and the secondary learning estimates based on the use of (1), (2), or (3), depending on data availability[21].

According to these estimates, in more than half of the countries, at least half of the students in secondary school are failing to achieve the minimum standard, and in approximately one-quarter of the countries – largely low-income nations – more than three out of four are failing to reach this standard.

[20] See notes above for definitions of variables

[21] A full listing of the countries for which the different approaches were used is provided in the appendix.

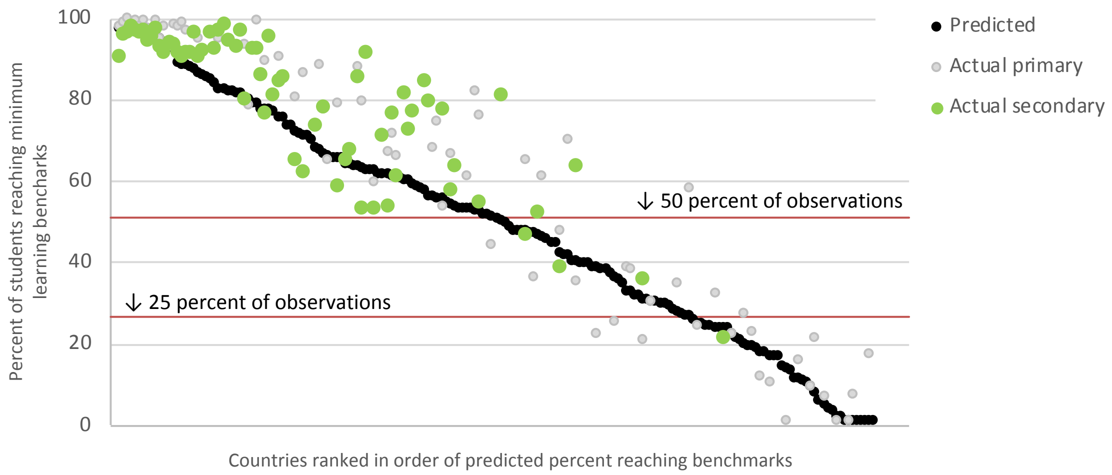


Figure 2. Country-level observed and predicted probability of reaching minimum levels of learning in secondary school (primary shown for comparison) with countries ranked in order of predicted values for 184 countries. Observed values based on PISA and TIMMS scores; predicted values based on Models 1, 2, or 3 from table 1 depending on available data. Source: Authors' calculations.

Conclusion

In closing, we note that in many developing countries, there is an absence of learning measures that can be standardized to a common international measure. However, learning measures are important to evaluate how well a country's education system is performing relative to peers and aspirational levels. This is necessary in order to identify countries that need significant improvement, and, to identify successful outliers that can be emulated, with the ultimate goal being to raise learning levels for all children in all countries of the world.

Currently, the UNESCO Institute of Statistics is working to fill this gap, using estimates and promoting the use of more learning assessments. In this paper, we have proposed two additional approaches currently not being used. One is a method for standardizing learning outcomes across multiple assessments that is simpler than what is currently used. The second is a method for estimating learning outcomes in countries without learning measures that can be used for standardization.

Both of these methods can contribute to the estimation of learning levels in developing countries. These estimates are a reasonable approach given the variability and the scarcity of available data, but clearly better, and more direct, measures of learning are needed in order to meaningfully pursue a goal of learning for all children.

References

[1] United Nations. Transforming our world: The 2030 Agenda for Sustainable Development. Adopted September 2015. New York, 2015. Available at: https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf. Accessed October 25, 2017.
[2] Education Commission. The learning generation: investing in education for a changing world. Report by the International Commission on Financing Global Education Opportunity. New York, 2016. Available at: http://report.educationcommission.org/wp-content/uploads/2016/09/Learning_Generation_Full_Report.pdf. Accessed October 25, 2017.
[3] Sheehan P, Sweeny K, Rasmussen B, et al. Building the foundations for sustainable development: A case for global investment in the capabilities of adolescents. The Lancet 2017; online April 19. Available at: http://dx.doi.org/10.1016/S0140-6736(17)30872-3. Accessed November 20, 2017.
[4] Education Policy and Data Center. Educating the world's children: patterns of growth and inequality. Washington DC, 2005.
[5] Education Policy and Data Center. Out of school children: data challenges in measuring access to education. Washington DC, 2013.
[6] UIS, UNICEF. Fixing the broken promise of education for all: findings from the Global Initiative on Out-of-School Children. Montreal: UNESCO Institute for Statistics (UIS), 2015. Available at: http://dx.doi.org/10.15220/978-92-9189-161-0-en. Accessed October 25, 2017.
[7] Global Partnership for Education. Results for learning report 2012. Washington DC, 2012.
[8] Crouch L, Merseth KA. Stumbling at the first step: efficiency implications of poor performance in the foundational first five years. Prospects; 2017. Online at https://doi.org/10.1007/s11125-017-9401-1.
[9] UNESCO. Pricing the right to education: the cost of reaching new targets by 2030. Policy Paper No. 18. Paris, 2015. Available at: http://en.unesco.org/gem-report/node/819. Accessed October 25, 2017.
[10] Wils A. The learning generation analytical background paper: trends, projections, education impacts. The Education Commission, New York, expected in 2018.
[11] Nguyen MC, Wodon Q. Child marriage and education: a major challenge. Draft. Washington DC: World Bank, 2012a. Available at: http://www.ungei.org/files/Child_Marriage_Edu_Note.pdf. Accessed November 20, 2017.
[12] Nguyen MC, Wodon Q. Child marriage, pregnancies and the gender gap in education attainment: an analysis based on the reasons for dropping out of school. Mimeo. Washington DC: World Bank, 2012b.
[13] Hattori H. Demographic and socioeconomic determinants of school attendance: an analysis of household survey data. Paper commissioned for Fixing the Broken Promise of Education for All: Findings from the Global Initiative on Out-of-School Children (UIS/UNICEF, 2015). Montreal: UNESCO Institute for Statistics (UIS), 2014. http://allinschool.org/wp-content/uploads/2015/01/OOSC-2014-Socioeconomic-determinants-final.pdf
[14] Wils A. Education, marriage and adolescent welfare: challenges, programs and projections. Supporting Paper to the UNFPA Project UNFPAUSA/14/037 Investment in Improving Health, Education and other Outcomes for Adolescents. Melbourne: VISES, Victoria University, 2015. Available at: http://vises.org.au/documents/2015_Wils_Education_Marriage_and_Adolescent_Welfare.pdf. Accessed October 25, 2017.
[15] Kan S, Wils A. Analysis of risk factors in education in developing countries: technical report. Background paper to The Learning Generation. New York: International Commission on Financing Global Education Opportunity, 2016. Available at: http://report.educationcommission.org/wp-content/uploads/dlm_uploads/2016/12/Analysis-of-Risk-Factors-in-Education.pdf. Accessed October 25, 2017.
[16] Hanushek EA, Woessmann L. The economics of international differences in educational achievement. NBER Working Paper No. 15949. Cambridge, Mass.: National Bureau of Economic Research, 2010.
[17] Barro R, Sala-i-Martin X. Economic growth, 2nd edn. Cambridge, MA: MIT Press, 2004.
[18] Patton GC, Sawyer SM, Santelli JS, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016; 387: 2423–78.
[19] Hanushek EA, Kimko DD. Schooling, labor force quality, and the growth of nations. American Economic Review 2000;90(5):1184–208.
[20] Hanushek EA, Woessmann L. Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. NBER Working Paper No. 14633. Cambridge, Mass.: National Bureau of Economic Research, 2009.
[21] Altinok N. A New international database on the distribution of student achievement. Background paper prepared for the Education for All Global Monitoring Report 2012 Youth and Skills: Putting Education to Work. Paris: UNESCO, 2012.
[22] Altinok N, Diebolt C, Demeulemeester J-L. A new international database on education quality: 1965–2010. Applied Economics 2014;46(11):1212–47.
[23] Angrist N, Patrinos HA, Schlotter M. An expansion of global data set on educational quality: a focus on achievement in developing countries. Policy Research Working Paper Series 6536. Washington DC: World Bank, 2013. Available at: http://datatopics.worldbank.org/Education/files/GlobalAchievement/Full.pdf. Accessed October 25, 2017.
[24] UNESCO. EFA global monitoring report 2008: education for all by 2015, will we make it? Paris, 2007. Available at: http://unesdoc.unesco.org/images/0015/001547/154743e.pdf. Accessed October 25, 2017.

[25] Hanushek EA. Interpreting recent research on schooling in developing countries. World Bank Research Observer 1995;10(2):227–46.
[26] Hanushek EA. Assessing the effects of school resources on student performance: an update. Educational Evaluation and Policy Analysis 1997;19(2):141–64.
[27] Hanushek EA, Luque JA. Efficiency and equity in schools around the world. Economics of Education Review2003;22:481–502.
[28] Wils A. Estimations for learning outcomes in low- and middle-income countries: technical report. Melbourne: VISES, Victoria University, 2017.
[29] UNESCO. EFA global monitoring report 2005: the quality imperative. Paris, 2004. Available at: http://unesdoc.unesco.org/images/0013/001373/137333e.pdf. Accessed October 25, 2017.
[30] UNICEF. Simulations for equity in education (SEE). New York, 2015. Available at: http://www.unicef.org/education/bege_SEE.html. Accessed May 19, 2016.
[31] Conn K. Identifying effective education interventions in Sub-Saharan Africa: a meta-analysis of rigorous impact evaluations. PhD Dissertation. New York: Columbia University, 2014.
[32] McEwan PJ. Cost-effectiveness analysis of education and health interventions in developing countries. J Dev Eff 2012;4:189–213.
[33] Snilstveit B, Stevenson J, Phillips D, NatCen, MV, Gallagher E, Schmidt T., . . . Eyers J. Interventions for improving learning outcomes and access to education in low- and middle-income countries: a systematic review. London: International Initiative for Impact Evaluation, 2015. Available at: http://www.3ieimpact.org/en/evidence/systematic-reviews/details/259/
[34] Kremer M, Brannen C, Glennerster R. The challenge of education and learning in the developing world. Science 2013;340(6130):297–300.
[35] Krishnaratne S, White H, Carpenter E. Quality education for all children? What works in education in developing countries. Working Paper 20. New Delhi: 3IE, 2013.
[36] Murnane RJ, Ganimian AJ. Improving educational outcomes in developing countries: lessons from rigorous evaluations. NBER Working Paper No. 20284. Cambridge, Mass.: National Bureau of Economic Research, 2014.
[37] Evans DK, Popova A. What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. Policy Research Working Paper 7203. Washington DC: World Bank, 2015.
[38] Conn K. The effectiveness of educations programs worldwide: evidence from a meta-analytic dataset. Background paper for the Education Commission. Washington DC, 2016. Available http://report.educationcommission.org/resources/ under tab "Performance" accessed October 25, 2017
[39] Blanton E, Ombeki S, Oluoch GO, et al. Evaluation of the role of school children in the promotion of point-of-use water treatment and handwashing in schools and households–Nyanza Province, Western Kenya, 2007. Am J Trop Med Hyg 2010;82:664–71.
[40] Bowen A, Ma H, Ou J, et al. A cluster-randomized controlled trial evaluating the effect of a handwashing-promotion program in Chinese primary schools. Am J Trop Med Hyg 2007;76:1166–73.
[41] Naz A, Khan W, Daraz U, et al. The impacts of corporal punishment on students' academic performance/career and personality development up-to secondary level education in Khyber Pakhtunkhwa Pakistan. Int J Bus Soc Sci 2011;2:130–40.
[42] Erulkar A, Medhin G, Weissman E. The impact and cost of child marriage prevention in three African settings. New York: Population Council, 2017. Available at: http://www.popcouncil.org/uploads/pdfs/2017PGY_ChildMarriageAfrica_report.pdf
[43] Hassana Alidou, Aliou Boly, Birgit Brock-Utne, Yaya Satina Diallo, Kathleen Heugh, H. Ekkehard Wolff. Optimizing Learning and Education in Africa – the Language Factor A Stock-taking Research on Mother Tongue and Bilingual Education in Sub-Saharan Africa. Libreville, Gabon: ADEA 2006 Biennial Meeting March 27-31, 2006. Available at: http://www.adeanet.org/adea/downloadcenter/Ouga/B3_1_MTBLE_en.pdf
[44] Wils A. Technical aspects of the education model. Supporting Paper to the UNFPA Project UNFPAUSA/14/037 Investment in Improving Health, Education and other Outcomes for Adolescents. Melbourne: VISES, Victoria University, 2016. Available at: http://vises.org.au/documents/2016_Wils_Technical_Aspects_of_Education_Model.pdf
[45] Sheehan P, Shi H. Enhanced educational outcomes, employment and productivity in developing countries. Paper under revision for Journal of Adolescent Health, 2018.

Table A1. Step-wise regression results for the learning production function.

Specifications	(0)	(1)	(2)	(3)	(4)
Ln GDP per capita	9.50***	9.09***	6.19***	9.88***	10.24***
Extraction rent as a percent of GDP		-0.81***	-0.68***	-0.69***	-0.69***
Average years of schooling adults 25-34			3.19***	1.79*	2.33***
Is or was socialist				14.3***	12.47***
Is Vietnam					20.61***
Constant	-11.95	-2.7	-9.17	-48.14***	-55.65***
N	74	74	74	74	74
R squared	0.32	.56	0.61	0.71	0.78

*** Significant at 1 percent; ** at 5 percent; * at 10 percent level.
Source: Commission analysis. Found in: “Vision access trends 8-25” tab “Learning model”.