

Learning Data for Better Policy: A Global Agenda

Nancy Birdsall, Barbara Bruns, and Janeen Madan

Abstract

Behind the learning crisis in much of the developing world is a huge data gap. Only a few middle income developing countries have the political incentives and technical capacity to develop and sustain national systems that measure what children are learning in school; most school children in the developing world have never taken a test that can be compared year over year or globally benchmarked. As enrollment has increased rapidly over the last two decades, policymakers and citizens have had no basis to assess whether more schooling has led to more learning, or respond with reforms and adjustments to improve education systems.

This paper sets out five concrete recommendations, addressed to the international community, which together could go a long way toward filling the global data gap on learning outcomes in the next decade. These recommendations constitute an ambitious agenda for developing countries and the broader development community—but one that is entirely affordable, costing only about \$400 million over the next 10-15 years.

Nancy Birdsall, Barbara Bruns, Janeen Madan. 2016. “Learning Data for Better Policy: A Global Agenda” CGD Policy Paper. Washington, DC: Center for Global Development.
<http://www.cgdev.org/publication/learning-data-better-policy>

The authors are grateful for the support and inputs of CGD colleagues William Savedoff and Justin Sandefur, and the full engagement and excellent contributions of members of a study group chaired by Nancy Birdsall, whose ideas and comments are at the heart of the recommendations outlined in this paper. The study group does not bear any responsibility for errors of fact or tone, or for the specific recommendations. This CGD policy paper is based on a report prepared for the International Commission on Financing Global Education Opportunities as background for their report, *The Learning Generation: Investing in Education for a Changing World*. The views and opinions in this background paper are those of the authors and are not endorsed by the Education Commission or its members.

Center for Global Development
2055 L Street NW
Fifth Floor
Washington DC 20036
202-416-4000
www.cgdev.org

This work is made available under
the terms of the Creative Commons
Attribution-NonCommercial 4.0
license.

Study group members

Nancy	Birdsall	CGD
Barbara	Bruns	CGD
Newman	Burdett	Cambridge University
Marguerite	Clarke	World Bank
Luis	Crouch	RTI
Deon	Filmer	World Bank
Eugenio	Gonzalez	Interagency Education Research Initiative (IERI)
Dirk	Hastedt	IEA
Elizabeth	King	Brookings Institute
Janeen	Madan	CGD
Silvia	Montoya	UIS, UNESCO
Anit	Mukherjee	CGD
Justin	Sandefur	CGD
William	Savedoff	CGD
Andreas	Schleicher	OECD, PISA
Sergei	Soares	CGD
Michael	Ward	OECD, PISA for Development

Contents

Background and context	1
Objectives	3
Recommendations.....	4
1.Support countries to establish robust national assessment systems	4
2.Help countries benchmark themselves internationally	5
3. Develop and pilot, as a global public good, a test for 9-year-olds to fill the gap in international assessments at the early learning stage.....	
4. Develop standardized learning modules that can be added to existing household surveys.....	
5. Establish the supporting architecture needed to fill the global data gap on learning outcomes	10
Epilogue	11
Figures and tables	14
List of acronyms	17
References	18
Appendices	19
1. Overview of existing international initiatives.....	1
2. What grade are 9-year-olds in?	20
3. What are the determinants of learning?	21
4. What can we learn from household surveys that we cannot learn from school-based surveys?.....	22

Background and context

There has been tremendous progress in increasing school enrollment rates globally, but education quality has lagged. Spending on education is not producing learning; illiteracy and innumeracy, including among children nearing the end of primary school, remain widespread. The evidence from recent studies is sobering. Globally, 250 million children of primary school age are not learning basic skills, even though half of them (130 million) have spent at least four years in school (UNESCO 2014). In India, less than half of the children surveyed in grade 5 could read a grade 2 level story, one in four could not read a simple sentence, and only slightly more than half could do subtraction.¹ In Tanzania, one out of five teachers were absent on the day schools were visited, and only ten percent of schools had all teachers at work.²

This crisis of learning has given rise to an emerging global framework to measure learning within and across countries. Learning—the central theme of the Learning Metrics Task Force (LMTF) 2013 report—is now a global priority. Learning is also underscored in the Education 2030 Framework for Action, the Sustainable Development Goal for Education (SDG 4), and the Global Alliance to Monitor Learning established in 2016.³ Over the last ten years, new initiatives sponsored by citizens (ASER, UWEZO) and financed at the international level (GPE, READ and SABER at the World Bank, EGRA, EGMA, PISA, TIMSS, PIRLS) have brought greater attention to measuring learning, especially in developing countries. (Acronyms here and below are spelled out at the end of the paper.) As a result, a growing number of countries are participating in regional and international tests and/or conducting their own national assessments of learning.⁴ However, there are enormous gaps in the data on learning outcomes across developing countries.

International assessments and regional initiatives such as LLECE in Latin America (Latin American Laboratory for Assessment of the Quality of Education) or PASEC (Program for the Analysis of CONFEMEN Education Systems) and SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality) in Africa, cover relatively few developing countries; the majority of school children in the developing world are not tested at all.⁵ Moreover, while there is a range of data sources on education systems, services, and outcomes, these are not always linked or available to researchers, particularly in developing countries, in a way that is user-friendly for policy relevant studies. Where data on children's learning outcomes are available, they are not easily linked to data on household characteristics, or to information on communities and school systems. The array of international actors (such as UIS/UNESCO, World Bank, GPE, LMTF, IEA, OECD/PISA, ASER, PASEC, SAQMEC, etc.) supporting learning measurement is a

¹ Based on ASER data for 2008. See Pritchett 2013 for more on how schooling is not producing “education” for millions of children in developing countries.

² Based on UWEZO data for 2011. See UWEZO-Tanzania, 2011.

³ See <http://brook.gs/2dMhFd6>; <http://bit.ly/2d2DwOq> and <http://bit.ly/2e1U34B>.

⁴ An assessment system is “a group of policies, structures, practices, and tools for generating and using information on student learning and achievement.” See Clarke 2012.

⁵ Justin Sandefur, “The Case for Global Standardized Testing,” CGD blog, April 27, 2016 available at: <http://bit.ly/1UlwHVY>.

positive sign of interest and a source of new and competing ideas, but also suggests high transaction and other costs of effective coordination. Existing international and regional assessments rely on ad-hoc donor contributions and do not always attract sustained financing. Several initiatives have also gone dormant because energy and/or resources ran out.

Financial support for countries to measure learning, and use the resulting data to improve school systems and address other root causes of low learning remains modest. Estimates suggest that only three percent of ODA for education is spent on global public goods, such as data and research. For the health sector, this figure amounts to an estimated 20 percent of ODA (Schäferhoff et al. 2016; see figure 2).

Furthermore, relatively few developing countries, and almost no low income countries, have standardized (equated over time) national assessment systems to track learning and provide a feedback mechanism to national education policies and programs. While a larger share of middle income countries undertakes some form of assessment, relatively few measure learning in a way that is globally benchmarked. A 2015 study of the Global Partnership for Education's (GPE) 60 partner countries found that assessment systems were "established" in two countries, "under developed" in 15, and "nascent" in 35, with data missing for eight countries. Also rare across the developing world is the "infrastructure" of data collection, organization, analysis, and feedback to educators, parents, and communities—a key input to making school systems effective and efficient, and to addressing inequalities in learning by region and income within countries.

Behind the striking lack of data on learning in many developing countries is not only, or mostly, lack of financing in the broad sense. In fact, the cost of maintaining a national system is small relative to overall spending on public education.⁶ Two other factors matter: lack of political incentives to "know" when and where schools are failing, and in many countries, lack of technical capacity.

The result of these gaps and challenges is that much of the learning data currently being generated by national and regional assessments cannot be benchmarked internationally. The SDG indicator framework, completed earlier this year, recommended three specific points of measure to track global progress on learning outcomes: "Percentage of children/young people in i) grades 2/3; ii) at the end of primary; and iii) at the end of lower secondary, achieving at least a minimum proficiency level in (i) reading and (ii) mathematics."⁷ However, to date there are no agreed standards of proficiency and no agreed tests to ensure that countries' measures are comparable to each other and over time.

⁶ The annual cost of assessment per secondary student as a share of annual government expenditure per secondary student ranges from 0.3 percent in East Asia and the Pacific to 0.1 percent in sub-Saharan Africa. See UIS 2016.

⁷ This refers to indicator 4.1.1, under SDG 4. Each point of measurement has its own value-added: grade 2/3 provides an indication at an early stage in the school cycle of whether children are learning foundational skills; end of primary school provides data on cumulative learning through primary schooling; end of lower secondary tests the skills needed for entering the labor force that all young adults should master.

For countries, the lack of robust measures of student learning inhibits policy development and accountability for results (Pritchett et al. 2013). At the global level, it restricts the power of research to help improve learning outcomes, strengthen the evidence base for decision making, and shift global aid towards outcomes-based financing. Although some impressive NGO initiatives to apply household-based tests and non-standardized tests of early grade reading and math (EGRA and EGMA) have generated shocking data on poor learning outcomes, there is little evidence to date that these have had an impact on governments' education policies (Behrman and Birdsall 2015).

In contrast, there is some evidence that data on learning assessments that can be globally benchmarked lead to policy reforms at the national level. A 2008 stakeholder survey found that half of the countries participating in the OECD's PISA test of 15-year-olds reported their government had launched reforms as a consequence of their PISA results (Figazzolo 2006). Examples include Germany's decision to rethink early tracking of students into vocational streams; Brazil's focus on reducing repetition; Mexico's sweeping 2008 program to raise education quality; Peru's reforms to raise standards for teachers; and initiatives in Turkey, Hungary, Indonesia, Kazakhstan, and Jordan (Breakspear 2012 and 2014; Bruns, Evans and Luque 2012; Bruns and Luque 2014).

The contrast with the health sector is particularly instructive. The global community knows immensely more about the determinants of infant and child mortality than of basic literacy and numeracy. This is in large part because of donor financing of household surveys in developing countries—in which standardized data on health indicators are collected every few years along with information on household income, access to services, and other community and national factors likely to impact health outcomes. The Demographic and Health Surveys (DHS) and the Living Standards Measurement Study (LSMS) have generated a rich body of data on health outcomes, allowing countries to compare themselves to their neighbors and others, and provides a platform for in-country research on what policy and program changes work best.

It is true that infant mortality, child nutrition, and other health indicators are easier to observe and less complex to measure than learning or cognitive skills. But digitization and declining transport and other costs have made large-scale data collection easier and cheaper. The key now is for the global education community to agree on appropriate measures of learning at key ages and on tests that can be used across the wide swath of the developing world that remains without these critical data.

Objectives

Our recommendations are framed around two medium-term objectives:

First, all countries design and implement national assessment systems that are nationally-owned, adapted to the country context (curricula, teaching standards, cultural contexts, etc.), and ensure equivalence over time. It often makes sense for assessments initially to be sample-based, while school systems develop the implementation capacity to ensure the

integrity of test administration and results. However, it is eventually desirable to conduct census-based assessments. The latter generate the school-level feedback on learning progress that is essential for parents and communities to hold school directors and system officials accountable for results. They also enable policymakers to use evidence to address policy questions of inclusion and equity across different communities, regions, ethnicities, religions, and gender. Ideally, national assessments are implemented at regular intervals, produce reliable, good quality data that are comparable over time, and are made available to the public.

Second, all countries develop the capacity to generate learning results that can be benchmarked internationally for the three recommended SDG points of measurement. Countries can do this by: i) participating in regional and/or international tests that measure learning outcomes on a globally comparable scale; ii) incorporating linking items from regional and/or international tests into their national assessments; iii) relying on participation in regional and/or international assessments until national assessment capacity can be developed; and iv) adding standardized learning modules to household income/expenditure surveys. Data from all of these sources can help measure and track progress towards SDG 4. They can also serve as valid outcome measures for the proposed “Global Offer for Learning,” (Sayedoff 2016), or other results-based programs such as Cash on Delivery Aid agreements, Social Impact Bonds, and Development Impact Bonds.

Recommendations

The following five recommendations build on existing initiatives to fill gaps in the emerging global framework to measure learning. For each recommendation, we describe the specific gap to be filled, the value-added of our proposed recommendation, and existing models that the international community can build on to jump-start implementation. Together this set of recommendations constitutes a vital and ambitious agenda for developing country governments and the broader development community—but one which is entirely affordable, adding up to direct costs of about \$400 million over the next 10-15 years.⁸

1. Support countries to establish robust national assessment systems

Gap

Lack of political incentives, technical capacity, and financing mean that relatively few developing countries have national assessment systems that are standardized (e.g., equivalent over time), or the capacity to use national results for evidence-based policy and accountability-enhancing feedback to parents and communities.

⁸ See table 2 for details on the estimated cost of each recommendation. Note that the \$400 million total does not include the costs for recommendation #5 on a “Learning Initiative Facility.”

Recommendation 1 [\$200 million]⁹

Donors contribute US\$200 million over the next ten years to support developing countries wishing to establish robust national assessment systems, and develop the infrastructure to use the resulting data for policy.

An estimated 50 developing countries (primarily low income countries) could receive up to \$4 million each over a 3-5 year period. International donors' support for technical capacity building should respond to specific requests from countries ("demand-driven") wishing to set up new assessment systems or strengthen existing ones.

Value-added: Establishing a dedicated, adequate, and steady global source of funding would finance the technical "infrastructure" (training of technical staff, international or regional advisors, organization, item banks, test construction and equipment) needed to design and implement robust national assessments, and make the resulting data available and useable in aggregate form to the public and in disaggregated form (anonymized, public use databases) to national and international researchers for cross-country research on the impact of education policies and programs.

Model to build on: An expanded READ at the World Bank (with modest changes to current design, and a different acronym) supported by additional funders.

2. Help countries benchmark themselves internationally

Gap

Only a few developing countries participate in regional or international assessment programs that allow country progress to be benchmarked against other countries (see figure 1). SDG progress on learning cannot be measured without systematic, coordinated efforts at the global level to allow international benchmarking of assessments going forward. The proposal by UNESCO's Institute for Statistics (UIS) to foster "linking" among existing national, regional, and international assessments so as to permit comparisons across countries and over time, is in its infancy and thus far unfunded.

Recommendation 2A [\$150 million]

Provide technical and financial support to low and lower-middle income developing countries (approximately \$1 million each) opting to participate in regional and/or international assessments that are equivalent over time and globally benchmarked.¹⁰ These resources could be pooled under the auspices of the "Learning Initiative Facility" described below, for deployment on demand from countries. Although country demand would trigger this funding, it is understood that part of the allocation would cover operating costs of the assessment agencies, for test development, analysis, and technical support to participating countries.

⁹ See table 2 for details on the estimated cost of each recommendation.

¹⁰ Upper-middle income countries, many of which already participate in international assessments, would not be eligible; see table 2 for more details on the cost estimation.

Value-added: Participating in international and regional assessments builds national technical capacity through learning-by-doing. Experience shows that these exercises generate positive spillovers for establishing sound national assessment systems that are adapted to the country context. Furthermore, evidence suggests that once countries participate in international or regional assessments—such as PISA, TIMSS, PIRLS, LLECE, SACMEQ and PASEC—that provide assistance with sampling and implementation, as well as analyzing and using resulting data, they are likely to continue to participate in subsequent years.

Models to build on: A number of existing international and regional assessments test literacy and numeracy among children and young adults. Most of these target the middle or end of primary school and end of lower secondary school—roughly corresponding to the second and third of the three points of measurement outlined under SDG 4 (see table 1). Important exceptions are PASEC and LLECE, the francophone African and Latin American regional assessments, both of which measure learning in grades 2 and 3.

Recommendation 2B [\$20 million]

Develop reliable and valid items for cross-linking existing regional and international learning assessments. These resources could be managed by UIS.

Value-added: Efforts to develop “common constructs” for measuring literacy and numeracy contribute an important global public good. Data on literacy and numeracy outcomes that are globally comparable would benefit countries, as well as the international community. The first step would be for international and regional assessment bodies to coordinate with country stakeholders and donors to create a harmonized assessment framework and a set of anchor items reflecting common constructs for literacy and numeracy, appropriate for each of the three recommended SDG measurement points (Sandefur 2016a and 2016b). These items would be open source, and managed by an agreed gatekeeper to assure confidentiality and preserve quality by releasing items responsibly to testing bodies and researchers. The next step would be to help countries introduce anchor items into their regional and national assessments. This would enable all countries to benchmark their progress against global standards, a critical first step towards SDG target 4.1: “By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes.”

Models to build on: A number of existing regional and international assessments align with the points of measurement recommended by the SDG indicators. PISA for Development managed by OECD could serve as the recommended global test to measure learning among 15-year-olds (end of lower secondary school). Several existing assessments measure learning in mid-to-late primary school (TIMSS and PIRLS managed by IEA, as well as several regional assessments; see table 1), might be used or adapted for the second recommended point of measurement, the end of primary school.

3. Develop and pilot, as a global public good, a test for 9-year-olds to fill the gap in international assessments at the early learning stage

Gap

There is no existing international assessment that targets children in early primary school—the first point of measurement outlined in SDG 4 (see table 1 and figure 1). Two regional assessments—PASEC and LLECE—test children in grades 2 and 3, but currently they only cover about 12 percent of children globally.¹¹

Recommendation 4 [\$10 million]

Develop and pilot a test of basic literacy, numeracy, and critical thinking skills for 9-year-olds, primarily administered in schools across relevant grades (parallel to PISA for 15-year-olds) to fill the gap in assessments at the early learning stage.¹² In countries where a significant share of 9-year-olds is not in school, a survey application in a representative sample of households would be conducted, parallel to the approach used by PISA for Development. To create incentives for low and lower-middle income countries to participate despite the risk of embarrassing results, any country that applied the assessment could be eligible to receive payments under the proposed “Global Offer for Learning” (Sayedoff 2016). This approach would complement, existing international assessments such as PIRLS and TIMSS that are grade-based (which makes comparisons across countries less transparent), and test students at a later grade level. The “Learning Initiative Facility” (described in more detail below) would outsource development of this test to an organization with the necessary technical expertise.

Value-added: The most powerful evidence of the global learning crisis are studies from India and Sub-Saharan Africa documenting that 50-60 percent of children today arrive at the final year of primary school without minimally proficient levels of literacy and numeracy.¹³ Years in school without learning makes schooling investments—for individual children, for countries, and for donors—hugely inefficient. A signal to school systems of widespread failure in the early teaching of basic literacy and numeracy is needed to drive reform of unrealistic curricula, ineffective teacher preparation, late entry to schooling, and lack of support for children falling behind.

Available data suggests that in many low income countries, the highest share of 9-year-olds are in grade 2 and 3 (see appendix 2). Therefore, focusing on 9-year-olds corresponds to the 2nd/3rd grade point of measurement recommended for SDG monitoring. An age-based test

¹¹ Justin Sandefur, “The Case for Global Standardized Testing,” CGD blog, April 27, 2016 available at: <http://bit.ly/1UlwHVY>.

¹² An earlier version of this paper submitted to the Education Commission in May 2016, recommended a global test for 8-year-olds. The authors now recommend age 9 for three reasons: 1) age 9 corresponds to grades 2 and 3 in many low-income countries (as shown in appendix 2); 2) age 9 is the modal age of children taking the TIMSS and PIRLS (grade 4) tests, which would facilitate benchmarking with a large pool of OECD countries at no additional cost; and 3) the six year difference between a test for 9-year-olds and PISA for 15-year-olds is equivalent to two rounds of PISA, which offers the possibility of tracking cohorts over assessment cycles.

¹³ PASEC 2015. Education System Performance in Francophone Sub-Saharan Africa.

has two key advantages vis-a-vis a grade-based test at this level. First, it is by design independent of the curriculum and focused on the core skills that must be acquired within the first few years of schooling for any further time in school to have an impact. Second, it reveals the share of children not entering school on time and/or excessively repeating grades—additional important benchmarking data for policymakers and international donors. Some of the most important education reforms stimulated by countries' participation in PISA, for example in Brazil and Germany, derived from the recognition that the share of 15-year-olds in 9th grade was behind that of other countries due to excess repetition, in the case of Brazil, or streamed into vocational education earlier than in other countries, in the case of Germany. Further support for prioritizing learning measurement in the earliest years of school comes from research by James Heckman showing that returns to human capital investments are the highest for younger children, and returns decrease as age increases.¹⁴

Models to build on: The Latin American *Laboratorio* (LLECE, SERCE, TERCE) and PASEC in Africa do an excellent job of measuring early literacy and numeracy skills and should anchor the design of test items appropriate for 9-year-old children across a wider range of developing countries and language groups. Results of the new globally-applicable instrument aimed at 9-year-old children would remain on a scale that could be equated easily to PASEC and LLECE results for 9-year-olds in their samples, meaning that no country currently covered by an existing regional assessment would need to change. Making the application sample based, and following PISA/PISA for Development protocols for testing an age-based (rather than grade-based) sample of students across a representative sample of schools would ease administration. A protocol to complement the school-based application with a household sample, where primary school enrolment or attendance rates are low, is also recommended.¹⁵ The household-adapted version of the test could be used more broadly in household surveys (see recommendation 4).

4. Develop standardized learning modules that can be added to existing household surveys

Gap

The international community has not supported the collection of data on learning in the context of household surveys that would allow for analysis of the determinants of learning at the household, school, and community levels—an important complement to data collected in schools. Instead, students and scholars working on education, especially in developing countries, tend to collect these data through small-scale efforts that lack comparability and external validity, and increase costs. Datasets combining household, school, and system level information—that have never been systematically combined—could trigger a new generation of social science research on the determinants of learning. This work would support evidence-based policy and program changes within countries, and a new generation of ideas, products, and business models to transform school systems across the developing

¹⁴ More information on Heckman's research is here: <http://bit.ly/1dO7SZ8>.

¹⁵ The World Bank's Service Delivery Indicators show that enrolled children's attendance is often low.

world—marking a shift from the “industrial” model focused on inputs to a “smart data” model focused on learning outcomes.

Recommendation 4 [\$20 million]

Develop standardized learning modules to test 9- and 15-year-olds encountered in households sampled in national surveys (e.g., income and consumption surveys) and internationally-sponsored surveys. Examples include the World Bank-sponsored LSMS and the USAID-sponsored DHS, which have provided a platform for research within and across countries on how economic policy affects household livelihoods and health, respectively. That research has in turn provided the basis for country policy and programs decisions.

The learning modules should be drawn from, and thus equated with, the tests we recommend for 9- and 15-year-olds in national and international learning assessments. Harmonizing the contextual data collected from school-based and household-based assessments is also essential. Convergence on a subset of questions and concepts related to household, school, and community characteristics that could be used consistently by researchers across countries would increase data comparability and research impact. The “Learning Initiative Facility” would outsource development of these modules to an existing organization with the necessary technical expertise in learning metrics and experience in the application of nationally representative household surveys.

Value-added: In addition to literacy and numeracy assessments, contextual data on student, household, school, and system characteristics help to complete the feedback loop that enables policymakers to improve learning outcomes. A number of different sources of data contribute to understanding the determinants of learning: information on inputs (textbooks, teachers); political dynamics, policies and institutions; service delivery indicators; household behaviors and income, education system characteristics; and learning outcomes (see appendix 3).¹⁶ Education Management Information Systems and the World Bank’s Service Delivery Indicators, for example, generate information on inputs and service delivery. The World Bank’s SABER initiative guides the collection of comparable data on the policy, legal, and institutional frameworks that shape education systems.

Household surveys are an essential complement to administrative data sources; household data are the only source of insight into behaviors and characteristics that can fundamentally shape children’s school attendance and learning progress. However, in most developing countries the ability to link a child’s learning outcomes with data on the household in which she lives is missing. Adding learning modules to household surveys poses challenges; home-based assessments do not provide a controlled testing environment. However, these assessments can collect data on family background, income, incentives, and behaviors—factors influencing schooling demand—that are an invaluable complement to data on supply-side factors such as teachers and schools, gathered in school-based assessments.

¹⁶ This draws on a presentation by Deon Filmer at the Building Evidence in Education Working Group Meeting held in April 2016 in Washington, DC. See appendix 3.

Identifiers that link children in the household to the primary or secondary school they attend, whether public or private, will increase the utility of household data for addressing “systems” questions. As an additional value-added, household surveys also enable researchers to link information on learning with data on health (e.g., individual health data including nutritional status as well as access to health services), for example (see appendix 4).

A learning module administered through household surveys is especially important in developing countries where a large proportion of 15-year-olds is out of school.¹⁷ A school-based test will not capture a nationally representative sample of children. As out-of-school children are often the most marginalized, a world concerned with learning for all should not leave these children unrepresented in national and global measures.

Models to build on: A learning module for 9-year-olds should be drawn from, and equated with, the 9-year-old test in recommendation 3 (building on *Laboratorio* and PASEC tests). The module for 15-year-olds should use PISA for Development’s household-based test for out-of-school 15-year-olds, as well as draw on lessons from UNICEF’s experience with the Multi-Indicator Cluster Survey (MICS).

5. Establish the supporting architecture needed to fill the global data gap on learning outcomes

Gap

The current institutional landscape constitutes an uncoordinated welter of different national, regional, international, citizen-based, and research-specific assessments and tests—none of which is universal in coverage or directly comparable with the others. The sum of global spending on these independent efforts is non-negligible and the lack of coordination means large missed opportunities for national benchmarking and cross-country research on how to raise student learning. UIS has taken a lead role in bringing the agencies that manage these different assessments together for consultation purposes. By launching the Global Alliance to Monitor Learning in 2016, UIS has begun trying to coordinate the development of common constructs that could, over time, create a platform for comparability and global benchmarking. But neither UIS nor any other single institution at present has the fiduciary and legal capacity to raise, manage, and allocate funds to existing agencies, programs, and countries to advance a coordinated agenda.

¹⁷ For example, Vietnam, Indonesia, Jordan, and Lebanon have a high out-of-school rate (exceeding 10 percent) among lower secondary school-age children. See Bloem 2013.

Recommendation 5

Build on the UIS Global Alliance to Monitor Learning to create a 10-15 year “Learning Initiative Facility” (sunset in 2030), bringing together philanthropic, donor, civil society, governments, and other groups with a core mission to improve country and global data on children’s learning, and promote use of these data to inform education and other policies, practices, and programs in developing countries.

Value-added: The “Learning Initiative Facility” would support developing countries to conduct their own national assessments to an appropriate standard, build analytical capacity for disseminating and using results, and support participation in international and regional assessments. The core funding base should be countries themselves, but the Initiative would leverage resources from international actors (multilateral development banks, bilateral donors, global philanthropies). The Initiative would serve as an honest broker in raising and deploying resources to developing countries to improve learning measurement.

Governance: The “Learning Initiative Facility” would be a non-profit entity with technical, fiduciary, and legal capacity to receive and allocate resources to agencies, programs, and countries to address the four preceding recommendations. Its work would be concentrated over about ten years, and it would close its doors no later than 2030. A small staff would work under the auspices of a governing body (a “board”), comprised of individuals with a diverse mix of the skills, perspectives, and experiences required by the organization. The board would be elected by funders to renewable three-year terms but would not directly represent any particular funding organization. Technical advisory groups drawn from existing assessment and technical bodies (OECD, IEA, SACMEQ, PASEC, LLECE, READ, REACH, RTI) would support the “Learning Initiative Facility” in the development of strategy, program oversight, and other technical areas. However, as they are expected to be major grant recipients under the Initiative, they would have no legal and fiduciary responsibility or representation in the Initiative's governance or any of its financing decisions.

Epilogue

The Education Commission report released in September 2016 powerfully depicted the global learning crisis. There is evidence that only half of primary school-aged children in low and middle income countries and little more than a quarter of secondary school-aged children are mastering basic primary- and secondary-level skills.

The Commission calls for significant increases in annual financing for education by 2030—from an estimated \$1 to \$2.7 trillion in public expenditures by low and middle income countries themselves, and from \$16 to \$89 billion by the international community (see figure 33, p. 105 in the Commission’s report). But it also recognizes that more spending will result in more learning only if countries undertake reform of education systems with a focus on delivering measurable increases in learning. For changes in school systems and other policies that affect educational progress to be effective, reforms have to be shaped by evidence. But without sound measures of student learning that are comparable over time, policymakers

lack the most basic tool for designing reforms and programs, evaluating impact, and improving the effectiveness of their own—and donors’—education spending. In the absence of robust data on student learning, the increases in investment recommended by the Commission risk yielding only limited returns in learning, and in the long-run skills development critical for sustainable growth across the developing world.

Specific recommendations of this paper endorsed in the Commission’s report include:

- **All “countries should develop their own national student assessment systems”** (p. 53). This corresponds with our recommendation #1. Our paper proposes \$200 million in direct support to developing countries to strengthen their national data systems for measuring student learning that will support program and policy evaluation, and orient system-level reforms. In addition, our paper proposes \$150 million to support countries to participate in international and regional assessments (recommendation #2A) and \$20 million to support UIS work on “linking items” that can build international benchmarking into national assessments (recommendation 2B).
- The international community should define a **“lead global learning indicator...based on the learning and skills expected by a given age”** (p. 56 of the Commission’s report). This corresponds with our recommendation #3. We suggest, in particular, that 9-year-olds be tested, independent of the grade they are in. Developing a global age-based test for 9-year-olds would help fill an important data gap: the lack of globally comparable data on learning outcomes in the first years of primary school. Our recommendation calls on donors to spend \$10 million to develop and pilot a test for 9-year-olds. If this instrument is also incorporated into a learning module that can be added to existing household surveys, data combining household, school, and community-level information could support a new generation of research on the determinants of learning. Our recommendation #4 is that donors should spend \$20 million to develop these learning modules.
- The **“financial, technical and capacity-building support of global partners should be harnessed through a new Global Education Data Initiative”...“building on and expanding the UNESCO Institute for Statistics and the new Global Alliance to Monitor Learning”**(p. 57 of the Commission’s report). This corresponds to our recommendation #5 proposing a 10-15 year “Learning Initiative Facility” (sunset in 2030) to support the Global Alliance to Monitor Learning. In addition, our paper proposes a specific governance structure that would ensure this entity has the technical, fiduciary, and legal capacity to advance a coordinated agenda. Our emphasis that the Initiative would have a core funding base from countries themselves while also leveraging resources from international actors is also closely aligned with the Commission’s focus on increased domestic spending on education.

While the Commission does not make a specific recommendation on financing for global education data, it calls for more spending in this area, noting that only three percent of education ODA is currently spent on global public goods, compared with an estimated 20 percent of ODA for these areas in health. Our paper estimates that as little as \$40 million per year over the next ten years in incremental ODA focused on improving data on student learning could have substantial impact on countries' ability to implement the Commission's overriding vision of "reforming education systems to deliver results".¹⁸ This appears to be an investment well-worth making.

¹⁸ The implementation of our proposed recommendations stretches over 10-15 years, but spending needs to be launched over the next 10 years.

Figures and tables

Figure 1: Whose Learning Gets Measured?

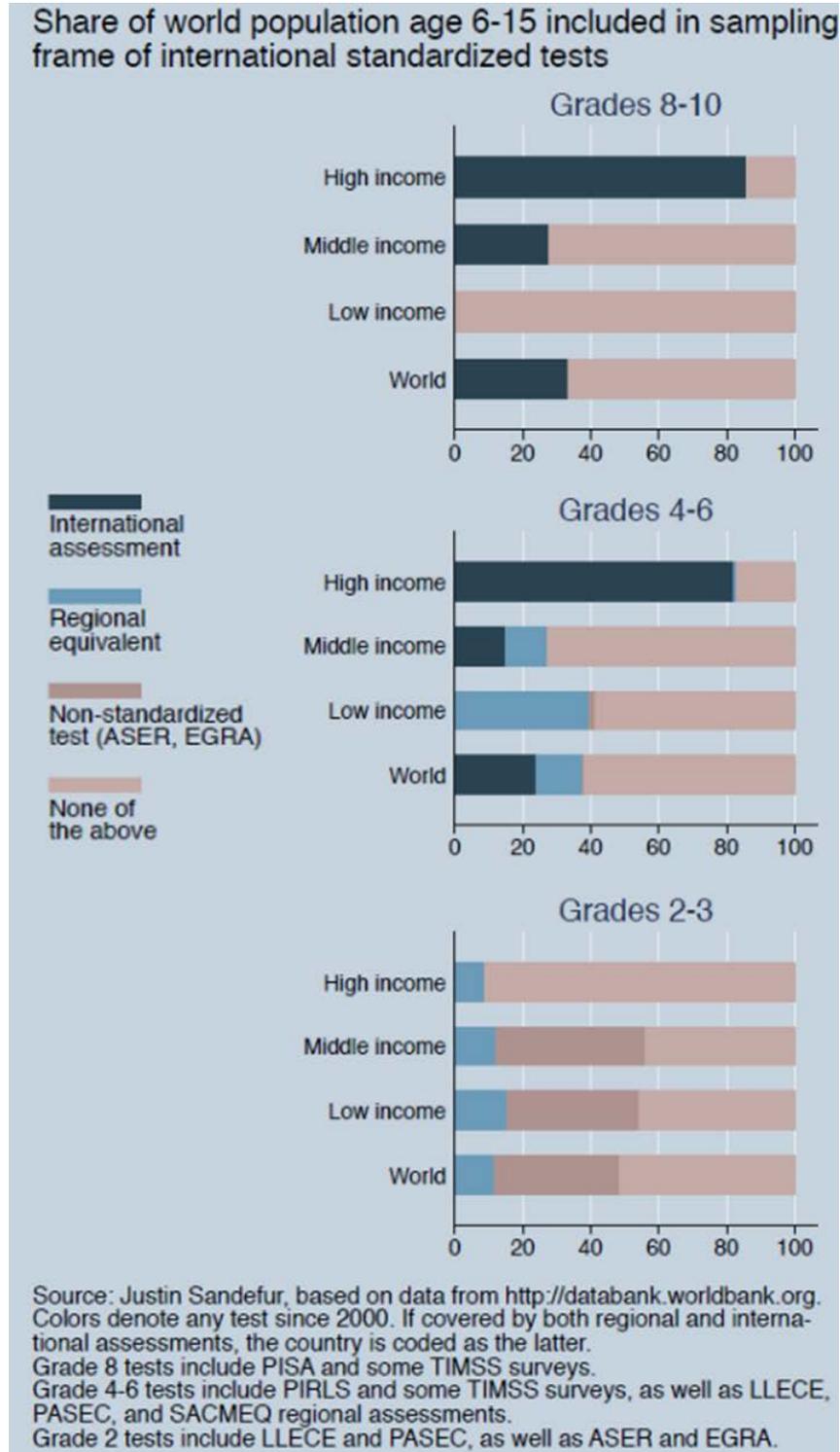
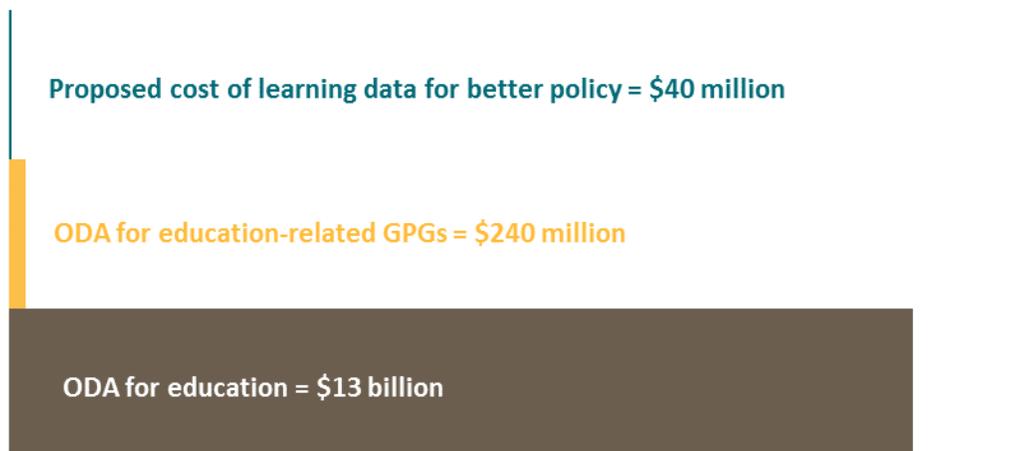


Figure 2: Financing for education (annual costs in US dollars)



Notes: ODA for education from OECD-DAC for 2015; ODA for education-related global public goods is based on Schäferhoff et al. 2016, as cited in the Education Commission report; proposed annual costs are based on recommendations 1 through 4 of this paper (see breakdown of costs by recommendation in table 2). For the proposed annual costs, note that while implementation of our proposed recommendations 1 through 4 stretches over 10-15 years, spending will need to be launched over the next 10 years. Total annual public expenditure for education by all low and middle income countries is estimated at \$1 trillion (see the Education Commission report).

Table 1: Regional and international assessments by grade

Grade	Assessment	# of countries	Reading	Numeracy and mathematics
2	PASEC	10	X	X
3	TERCE	15	X	X
4	PIRLS	48	X	
	PILNA	13	X	X
	TIMSS	63		X
5	N/A	N/A	N/A	N/A
6	TERCE	15	X	X
	PASEC	10	X	X
	PILNA	13	X	X
	SACMEQ	15	X	X
7	N/A	N/A	N/A	N/A
8	TIMSS	63		X
9 or 10	PISA (15-year-olds)	65	X	X

Source: Institute for Statistics, UNESCO.

Table 2: Rough cost estimates by recommendation¹⁹

Recommendation	Cost details	Total
1. Technical support to establish robust national assessment systems	Up to \$4 million over 3-5 years per country, for 50 countries ²⁰	\$ 200 million
2A. Support for national participation in international and regional assessments	Approximately \$1 million per assessment for 3 assessment cycles in 50 countries ²¹	\$150 million
2B. Cross-linking learning assessments	-	\$20 million
3. Develop and pilot a test for 9-year-olds		\$10 million
4. Develop learning modules to be added to household surveys		\$20 million
TOTAL		\$400 million

¹⁹ The implementation of our proposed recommendations stretches over 10-15 years, but spending needs to be launched over the next 10 years.

²⁰ Based on estimates from READ.

²¹ Based on estimated costs of international assessments (including capacity building support), such as PISA, PIRLS, and TIMSS.

List of acronyms

A4L	Assessment for Learning
ASER	Annual Status of Education Report
DHS	Demographic and Health Survey
EGMA	Early Grade Math Assessment
EGRA	Early Grade Reading Assessment
GPE	Global Partnership for Education
IEA	International Association for the Evaluation of Educational Achievement
LaNA	Learning and Numeracy Assessment
LLECE	Latin American Laboratory for Assessment of the Quality of Education
LMTF	Learning Metrics Task Force
LSMS	Living Standards Measurement Study
MICS	Multiple Indicator Cluster Survey
OECD	Organization for Economic Co-operation and Development
PASEC	Program for the Analysis of CONFEMEN Education Systems
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
READ	Russia Education Aid for Development
RTI	Research Triangle Institute
SABER	Systems Approach for Better Education Results
SACMEQ	Southern & Eastern Africa Consortium for Monitoring Educational Quality
SDG	Sustainable Development Goal
SDI	Service Delivery Indicators
TIMSS	Trends in International Mathematics and Science Study
UIS/UNESCO	Institute for Statistics of the United Nations Educational, Scientific and Cultural Organization

References

- Behrman, J. R. and Nancy Birdsall. 2015. "Information, Assessment and the Quality of Education around the World in a Changing Global Labor Market: Making More People Winners." Switzerland: Global Citizens Foundation.
- Bloem, Simone. 2013 "PISA in low and middle income countries." OECD Education Working Papers, No. 93. Paris, France: OECD Publishing.
- Breakspear, S. 2012. "The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance." OECD Education Working Papers, No. 71. Paris, France: OECD Publishing.
2014. "How Does PISA shape education policymaking? Why how we measure learning determines what counts in education." Melbourne, Australia: Centre for Strategic Education.
- Bruns, Barbara, David Evans, and Javier Luque. 2012. "Achieving World Class Education in Brazil: the next agenda." Washington, DC: World Bank.
- Bruns, Barbara and Javier Luque. 2015. "Great Teachers: How to Raise Student Learning in Latin America and the Caribbean." Washington, DC: World Bank.
- Clarke, Marguerite. 2012. "What Matters Most for Student Assessment Systems: A Framework Paper." Washington, DC: World Bank.
- FIGazzolo, L. 2007. "Impact of PISA 2006 on the Educational Policy Debate." Education International.
- International Commission for Financing Global Education Opportunity (Education Commission). 2016. "The Learning Generation: Investing in education for a changing world." New York, NY.
- Institute for Statistics. 2016. "The Cost of Not Assessing Learning Outcomes." Information Paper No. 26. Montreal, Canada: UNESCO Institute for Statistics (UIS).
- Learning Metrics Task Force (LMTF). 2013. "Toward Universal Learning: Recommendations from the Learning Metric Task Force." Montreal, Canada and Washington, DC: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.
- PASEC. 2015. "Education System Performance in Sub-Saharan Africa: Competencies and Learning Factors in Primary Education." PASEC (Programme d'Analyse des Systèmes Educatifs de la CONFEMEN). Dakar.
- Pritchett, Lant, Rukmini Banerji, and Charles Kenny. 2013. "Schooling Is Not Education! Using Assessment to Change the Politics of Non-Learning." A Report of the Center for Global Development Study Group on Measuring Learning Outcomes. Washington, DC: Center for Global Development (CGD).
- Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning.* Washington, DC: CGD.
- Sandefur, Justin. 2016a. "Linking Regional and International Assessments: Prospects for Creating a Global Learning Metric." Forthcoming CGD Working Paper. Washington, DC: CGD.
- Sandefur, Justin. 2016b. "Internationally Comparable Mathematics Scores for Thirteen African Countries." Forthcoming CGD Working Paper. Washington, DC: CGD.
- Savedoff, William. 2016. "A Global Offer for Learning: Based on Experiences with Paying for Results." Forthcoming CGD Policy Paper. Washington DC: CGD.
- Schäferhoff, Marco et al. 2016. "Rethinking the Financing and Architecture of Global Education." Prepared for the Education Commission. Berlin, Germany and Washington, DC: SEEK Development (SEEK) and Results for Development (R4D).
- UNESCO. 2014. "EFA Global Monitoring Report 2013-2014." Paris, France: UNESCO Publishing.
- UWEZO-Tanzania. 2011. "Are Our Children Learning? Annual Learning Assessment Report." Dar-es-Salaam, Tanzania: UWEZO.

Appendices

1. Overview of existing international initiatives

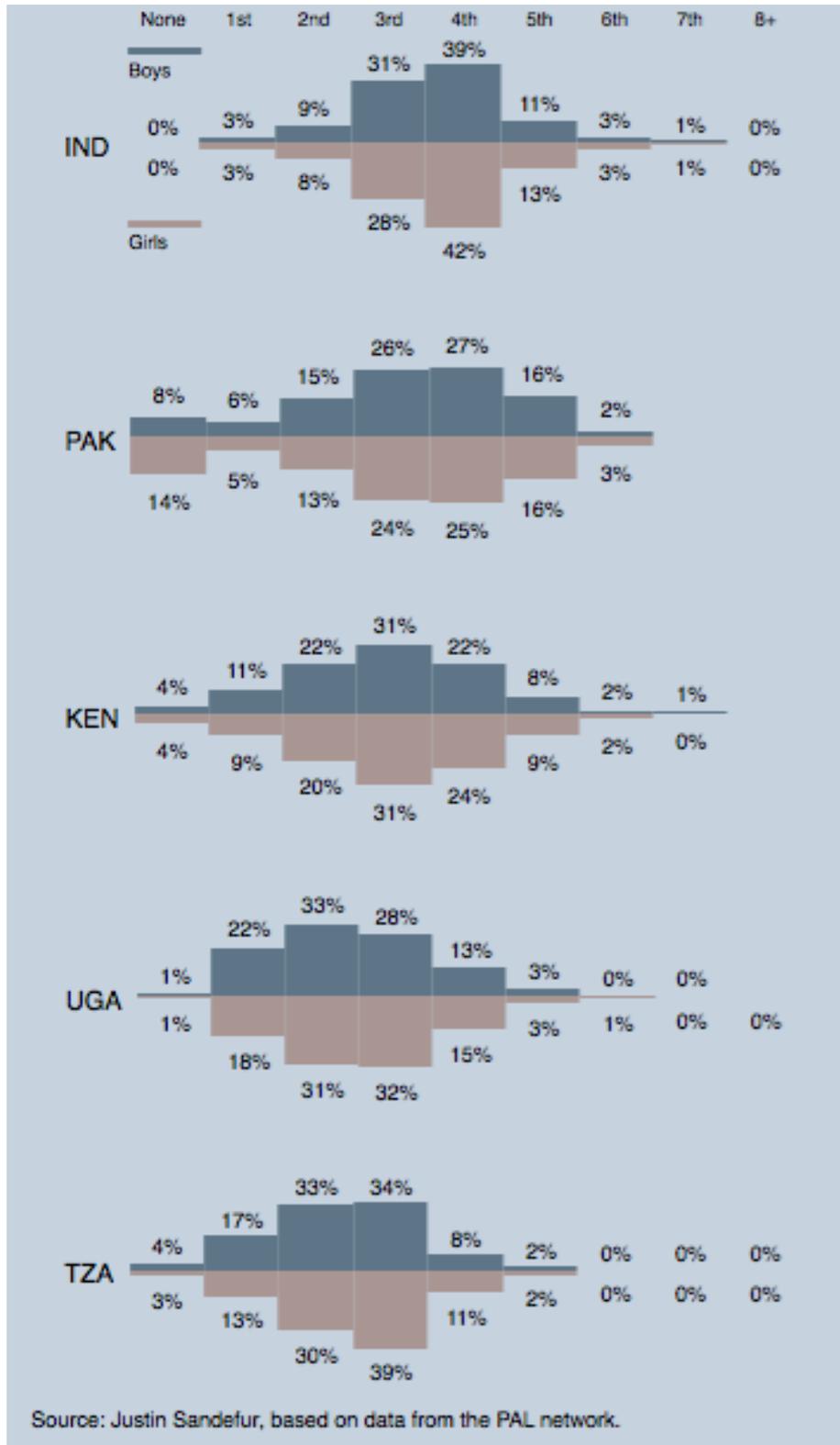
Over the past few years, a number of international initiatives have been launched to address the gap in data on learning outcomes. The following list provides an overview:

1. GPE, established in 2002, supports 60 developing countries to improve learning outcomes by funding components of education sector plans.
2. The LMTF helped shift the paradigm from access to access *plus* learning, and identified seven domains of learning.²²
3. UIS recently launched the Global Alliance to Monitor Learning—a global coordination platform of ongoing efforts to generate globally comparable learning data—and is also compiling a Catalogue of Learning Assessments (national and international) at primary and lower secondary levels.
4. The World Bank manages the READ Trust Fund, which has been supporting eight developing countries²³ to establish and/or strengthen assessment systems since 2008; and the World Bank SABER initiative’s diagnostic tools to assess education sector strengths and weaknesses includes a module on national assessment capacity.
5. Assessment for Learning (A4L)—a concept that originated from the LMTF recommendations and still in the preliminary phase— would serve as an international platform to coordinate technical support for national assessment systems.
6. PISA for Development is working with seven developing countries on an assessment for 15-year-olds that is benchmarked to the main PISA exam but appropriate for less developed countries; it will also apply the test in a sample of households in countries where a large share of 15-year-olds are out-of-school.
7. IEA is developing LaNA, a test of basic literacy and numeracy for grades 5 or 6 students that is benchmarked to TIMSS and PIRLS, but appropriate for developing countries.
8. UNICEF plans to integrate a short learning assessment of early reading and mathematics skills for 7-14-year-olds into MICS.

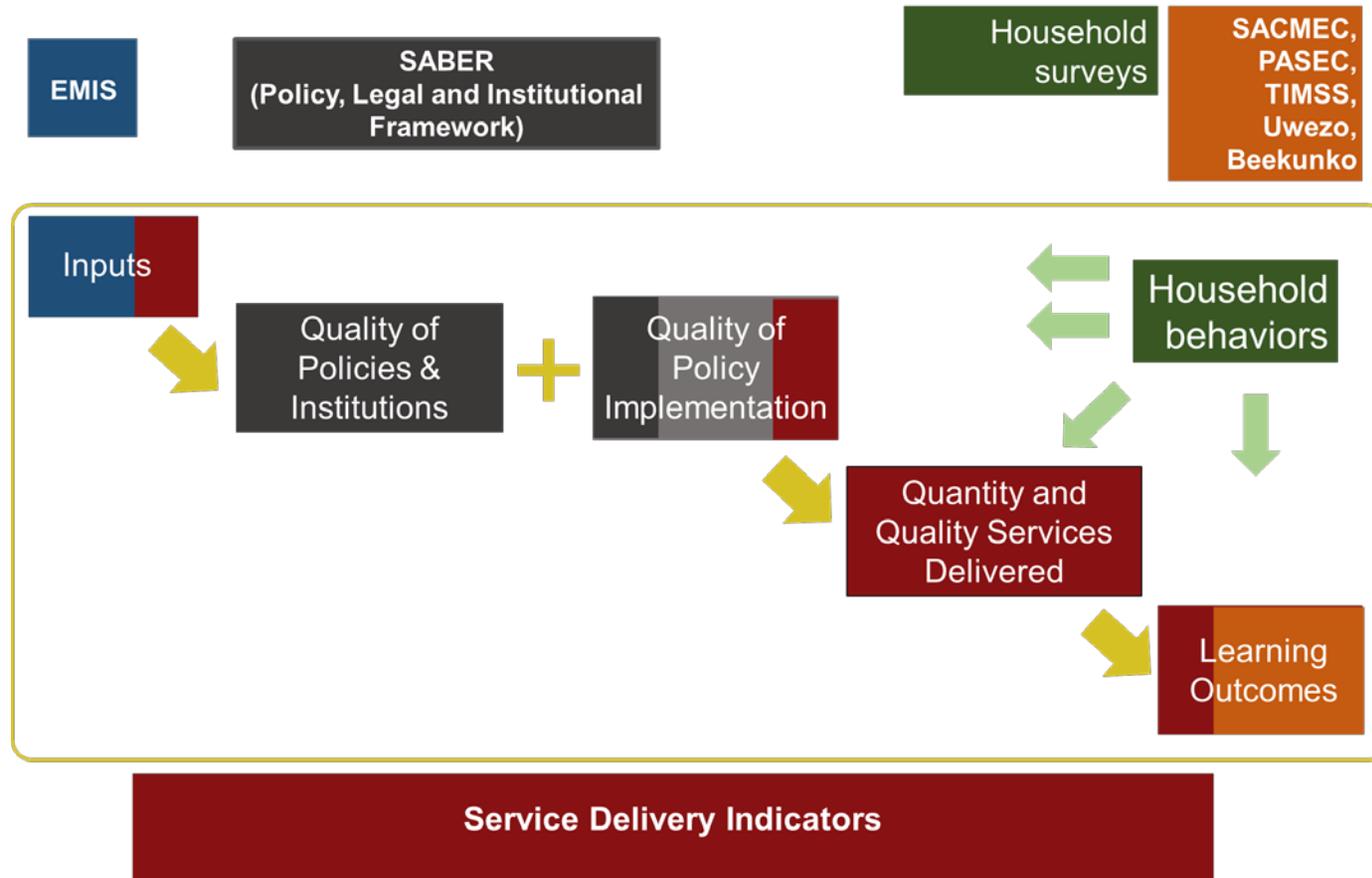
²² The seven domains include: science and technology, numeracy and mathematics, learning approaches and cognition, literacy and communication, culture and the arts, social and emotional, and physical and well-being.

²³ The eight countries supported by READ from 2008-2015 are: Angola, Armenia, Ethiopia, the Kyrgyz Republic, Mozambique, Tajikistan, Vietnam, and Zambia. READ just received funding for an additional three years, but the target countries for the next phase are not yet finalized.

2. What grade are 9-year-olds in?



3. What are the determinants of learning?



Source: Presentation by Deon Filmer at the Building Evidence in Education Working Group Meeting, April 2016, Washington, DC.

4. What can we learn from household surveys that we cannot learn from school-based surveys?

Learning assessments collect useful demographic and socio-economic information about students. A household survey would enable researchers to link data on learning with a rich array of data on other factors to better understand the determinants of learning and explore equity issues. Some examples include:

- Health: nutritional inputs and dietary information, health status (anthropometrics) and previous health conditions, health care of the child including pre- and neo-natal care, access to health services, etc.;
- Costs: school fees, other education-related expenditures;
- Transport and infrastructure: distance and time to nearest school, challenges affecting why children may not attend school; and
- Measures of school “systems”: private or public, whether unions exist or not.