MAY 2019

STRENGTHENING LEARNING ASSESSMENT SYSTEMS

A KNOWLEDGE AND INNOVATION EXCHANGE (KIX) DISCUSSION PAPER

GLOBAL PARTNERSHIP *for* EDUCATION

Acknowledgments

This paper was written by Kate Anderson, independent consultant and founder at Unbounded Associates, under the guidance of Ramya Vivekanandan, Senior Education Specialist and Thematic Lead for Learning Assessment Systems at the GPE Secretariat.

The inputs and feedback of the following individuals who reviewed this paper are gratefully acknowledged: Hassane Djibo Adamou, Ray Adams, Michelle Belisle, Jean-Marc Bernard, Chris Berry, Suman Bhattacharjea, Moritz Bilagher, Ryan Burgess, Manuel Cardoso, Esther Care, Talia de Chaisemartin, Marguerite Clarke, Luis Crouch, Dirk Hastedt, Sarah Hennell, Hilaire Hounkpodoté, Baela Raza Jamil, Luke Justice, Anil Kanjee, Joseph François Désiré Kauphy, Janeanne Kiviu, Asefa Leta, Silvia Montoya, Karen Mundy, Abdulkadir Yusuf Nuh, Eileen O'Malley, Folake Olatunji-David, Atilio Pizarro, Caroline Pontefract, Sara Ruto, Mioko Saito, François Sawadogo,Ursula Schwantner, Ernesto Trevino, Dan Wagner, Michael Ward and Camilla Woeldike.

In addition, we would like to acknowledge UNESCO Dakar (Secretariat of the Teaching and Learning Educators' Network for Transformation – TALENT) and UNESCO Bangkok (Secretariat of the Network on Education Quality Monitoring in the Asia-Pacific – NEQMAP) for allowing us to organize consultation sessions to gather feedback on this paper during the TALENT and NEQMAP regional workshops which were held in Dar es Salaam, Tanzania in July 2018 and in Penang, Malaysia in September 2018, respectively. Through these consultations, representatives from the following countries provided useful feedback on the paper: Australia, Bangladesh, Benin, Bhutan, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Comoros, Ethiopia, Fiji, The Gambia, Iran, Kyrgyzstan, Lao PDR, Lesotho, Liberia, Malawi, Malaysia, Maldives, Mali, Mongolia, Myanmar, Nepal, Niger, Papua New Guinea, Philippines, Senegal, South Sudan, Sri Lanka, Tanzania, Thailand, Timor Leste, Uganda, Uzbekistan, Vietnam and Zimbabwe.

Table of Contents

Executive Summary 6

1. Introduction 10

2. Paper development and consultation process 11

3. Challenges to building strong learning assessment systems 12

4. How GPE supports learning assessment systems 14

5. Global goods and innovation in learning assessment 16

6. Gaps in available global goods 24

7. Potential investment areas 24

Bibliography 31

Annex A. GPE fact sheet on learning assessments 35

Annex B. National and cross-national assessments in DCPs 2010–2017 38

A note on the KIX consultation process

The Global Partnership for Education's Knowledge and Innovation Exchange (KIX) thematic funding will support global and regional initiatives that use knowledge exchange, evidence and innovation to help developing countries solve critical educational challenges. It will support:

- **Capacity development and knowledge exchange among developing countries**: activities that strengthen national capacity through peer review and exchange; creation of learning modules and diagnostic tools, and face-to-face exchange
- **Evidence and evaluation**: activities that aim to consolidate and/or extend knowledge about how to improve educational outcomes and national education systems
- **Innovation pilots**: Piloting of approaches, methods, tools or products that solve persistent educational challenges

Investments will be guided by the priorities of developing country partners and allocated through a competitive process managed by an independent grant agent. Knowledge products, innovation pilots, and related tools developed through KIX funding will be shared through the Learning Exchange to amplify their uptake.

The purpose of this paper is to describe the current landscape in strengthening learning assessment systems and spark discussion and debate around potential areas for KIX investment. The paper is part of a series of discussion papers, drafted to support the engagement and consultation of developing country partners and technical experts in the initial design of the GPE Knowledge and Innovation Exchange. The ideas presented in the initial version of the paper served as a starting point for discussion and were modified significantly based on the consultation process, thereby resulting in this updated version. The consultations and revisions to the paper also informed a concept note that was developed to guide organizations wishing to access KIX funding, which will be made available in 2019.

Acronyms and Abbreviations

A4L	Assessment for Learning
ADEA	Association for the Development of Education in Africa
AEAA	Association for Educational Assessment in Africa
ANLAS	Analysis of National Learning Assessment Systems
ASER	Annual Status of Education Report
CLA	citizen-led assessment
DCP	developing country partner
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EMIS	education management and information systems
EQAP	Educational Quality and Assessment Programme
ESPIG	education sector program implementation grant
FCAC	fragile and conflict-affected countries
GAML	Global Alliance to Monitor Learning
GPE	Global Partnership for Education
GP-LA	Good Practices in Learning Assessment
GRA	Global and Regional Activities
IAEA	International Association for Educational Assessment
IEA	International Association for the Evaluation of Educational Achievement
IIEP	International Institute for Education Planning
KIX	Knowledge and Innovation Exchange
LACI	Learning Assessment Capacity Index
LaNA	Literacy and Numeracy Assessment
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
MICS	Multiple Indicator Cluster Survey
NALA	Network for African Learning Assessment
NGO	non-governmental organization
NCME	National Council on Measurement in Education
NEQMAP	Network on Education Quality Monitoring in the Asia-Pacific
OECD	Organisation for Economic Cooperation and Development
PASEC	Programme d'Analyse des Systèmes Educatifs de la CONFEMEN
PILNA	Pacific Islands Literacy and Numeracy Assessment
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PISA-D	PISA for Development
SABER	Systems Approach to Better Education Results
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SDG	Sustainable Development Goal
SEA-PLM	Southeast Asia Primary Learning Metric
SIDS	small island developing states
TALENT	Teaching and Learning: Educators' Network for Transformation
TERCE	Tercer Estudio Regional Comparativo y Explicativo
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Education, Cultural and Scientific Organization

Key Terms

Learning assessment. The process of gathering and evaluating information on what learners know, understand and can do.

Learning assessment system. The policies, institutions, assessment tools and practices that contribute to gathering and using information on how students are learning in the education system.

Classroom-based assessment. Assessments used to obtain evidence on the knowledge, skills and attitudes of learners for use by teachers to improve learning and teaching.

Examinations. Assessments undertaken to determine an individual student's mastery of specific knowledge and skills for the primary purpose of selection.

Large-scale assessments. These are system-level assessments for monitoring and providing policy makers and practitioners with information on the overall performance levels of education systems, changes in those levels, and related or contributing factors.

Sources: Clarke (2012), Wagner (2012).

Executive Summary

The importance of learning assessment systems

The effective use of assessment data is fundamental to education systems. Without data on how students are learning, it is impossible to know whether a country is succeeding in providing quality education to all its citizens. Learning assessments when analyzed with contextual data can be powerful tools for improving national education systems. A learning assessment system encompasses the policies, institutions, assessment tools and practices that contribute to gathering and using information on how students are learning in the education system.[1] It can include assessments designed and by teachers to inform their instruction, national examinations taken by all students, and sample-based assessments used primarily for policy making. A strong learning assessment system can help students, their parents, teachers, and school administrators track progress, and for intervening in cases where a child's learning is off track.

Learning outcomes are central to the global education agenda set out in the Sustainable Development Goals (SDGs) and the Education 2030 Framework for Action, so the measurement of these outcomes is also a global priority. SDG 4[2] lays out ambitious targets that include improving learning outcomes for all, from early childhood to adulthood. Half of the 10 education targets in SDG 4 relate to learning and skills.[3] Despite the benefits of monitoring learning outcomes, many countries do not have high-quality, relevant learning assessment systems. This is especially true in low-income countries, fragile and conflict-affected countries (FCACs), and small island developing states, many of which are Global Partnership for Education (GPE) developing country partners (DCPs).

Challenges to building strong learning assessment systems

Four main challenges to building a strong learning assessment system emerged from a desk review and consultations with DCPs and international experts: (1) quality of the assessment tools and the lack of technical expertise for assessment design, administration and analysis; (2) alignment across the various

[1] Clarke, Marguerite. "What Matters Most for Student Assessment Systems: A Framework Paper." Washington DC: World Bank, 2012.

[2] SDG 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

[3] SDG targets 4.1, 4.2, 4.4, 4.6 and 4.7, and their corresponding indicators, are all related to learning outcomes.

types of assessments used, and the positioning within national education systems; (3) assessments often not encompassing the most marginalized children, including those with disabilities and not in school; and (4) the lack of use of learning assessment results, which is influenced by how assessments are designed and disseminated, and the available expertise and resources for making change.

How the GPE supports learning assessment systems

Building strong learning assessment systems is a priority for GPE, as demonstrated by its results framework and funding model. The GPE 2020 Strategic Plan sets improved and more equitable learning outcomes as the first of its three goals. GPE's funding model also requires countries applying for a program implementation grant to either have a system in place to monitor learning outcomes or a plan to develop one. In 2017, the partnership's Board decided that if funding is insufficient to build these systems, then GPE funding should be used to finance these activities.

GPE's results framework monitors the percentage of DCPs that improved their learning outcomes in basic education during 2016–2020, and whether DCPs have a learning assessment system that meets quality standards. GPE country program implementation grants support 29 countries (71 percent of current grants) to develop various aspects of their learning assessment systems. To access GPE funding, all grantees must demonstrate that they have good quality data systems or a strategy for building them. As of 2016, only 32 percent of DCPs had learning assessment systems had one or the other.

To help DCPs meet these ambitious goals, GPE also supports regional and international activities that strengthen knowledge, evidence and national capacity. Previous grants have supported the UNESCO Institute for Statistics (UIS) to identify best practices in early reading assessment and a catalogue of learning assessments, and the Network on Education Quality Monitoring in the Asia-Pacific to build regional evidence and capacity. Following these activities, the Assessment for Learning (A4L) initiative was launched in July 2017 to build capacity for national learning assessment systems to measure and improve learning. A4L set the stage for KIX by developing a tool for countries to analyze their learning assessment systems and so contribute to efforts to strengthen them. A4L also provides support for regional networks that facilitate expertise for countries to build learning assessments that respond to national curricula and policy priorities. KIX grants will advance these efforts by further enabling knowledge exchange and supporting new global goods and building on the existing ones. These are outlined in the following section.

Global goods and innovations in learning assessment

Various tools and research studies are available to countries wishing to assess learning. International and regional comparative assessments provide valuable cross-national data on student learning across countries and often facilitate capacity building and peer learning opportunities.[4] Assessments of foundational literacy and numeracy skills, such as citizen-led assessments and early grade reading assessments, are not designed to be comparable across countries, but have a community of users to

[4] Including the Foundational Learning Skills Module of the UNICEF Multiple Indicator Cluster Survey (MICS), Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (PASEC), Pacific Islands Literacy and Numeracy Assessment (PILNA), Progress in International Reading Literacy Study (PIRLS), Programme for International Student Assessment (PISA and PISA-D), Southeast Asia Primary Learning Metric (SEA-PLM), Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), Tercer Estudio Regional Comparativo y Explicativo (TERCE), and Trends in International Mathematics and Science Study (TIMSS).

support countries in using these tools. A number of online platforms and professional organizations facilitate knowledge exchange, particularly for those working on large-scale assessments.[5]

There are also initiatives that support capacity building. These include books, guides and courses that can be used to build capacity for learning assessments that have been developed for low- and middle-income countries. For example, the Global Alliance to Monitor Learning (GAML), hosted by UIS, has guidelines on good practices in learning assessment and several "quick guides" on learning assessment issues for those implementing assessments.[6] The World Bank's Systems Approach for Better Education Results (SABER) Student Assessment module provides evidence-based rubrics for evaluating a country's learning assessment system, and has data from 38 countries. GPE is developing a complementary tool to support the diagnosis of learning assessment systems, the Analysis of National Learning Assessment Systems.

Existing global goods include innovative approaches to assessing learning and using the results to improve learning. UIS is refining the social moderation approach to equate proficiency levels for reading and mathematics across different countries and assessment tools. Nongovernment organizations are also using innovations to use learning assessment to teach to the ability levels of children[7] and to develop classroom assessments to measure 21st century skills.[8] Skills including problem solving and citizenship are also increasingly assessed in cross-national assessments.

Gaps in available global goods

Most global goods focus on three areas: (1) global goods that support groups of countries to design and administer specific large-scale learning assessments, (2) global goods that bring together networks of professionals working on assessment issues, and (3) online resources with general guidelines on how to develop quality learning assessments.

DCPs overwhelmingly said there is a gap in guidance for training teachers to conduct and use classroom-based assessments, and in good practices in examination reform. Although participation in large-scale assessments provides some capacity building for some ministerial officials, DCPs noted this needs to be more widespread to build an overall culture of evaluation. DCPs also noted inconsistencies across their learning assessment systems, and requested guidance on how to build strong institutions to oversee the entire national assessment portfolio. International experts and DCPs noted there is a lack of globally comparable data on learning outcomes, and they were interested in getting more countries to participate in cross-national assessments.

[5] Including most of the cross-national assessment agencies listed in footnote 3, plus UNESCO's IIEP Learning Portal, Global Alliance to Monitor Learning (GAML), International Association for Educational Assessment (IEA), National Council on Measurement in Education (NCME), Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP), Teaching and Learning: Educators' Network for Transformation (TALENT), Network for African Learning Assessment (NALA), and Association for Educational Assessment in Africa (AEAA).

[6] UIS (UNESCO Institute for Statistics) and ACER (Australian Council for Educational Research). *Principles of Good Practice in Learning Assessment*. Camberwell, Victoria: ACER, 2017.

[7] Banerji, Rukmini, and Michael Walton. "What Helps Children to Learn? Evaluation of Pratham's Read India Program in Bihar & Uttarakhand." New Delhi: Pratham, 2011; Save the Children. *Beyond School Walls: A Boost for Readers*. Fairfield CT: Save the Children, 2013.

[8] Care, Esther, Alvin Vista, and Helyn Kim. "Optimizing Assessment for All." Washington, DC: Brookings Institution, 2018.

Potential investment areas

Four main opportunities emerged from the consultation process for which global goods funded by KIX could be critical. These opportunities also align with the GPE results framework and funding model:

Global goods that support national learning assessment institutions. These include ministries of education and autonomous and semi-autonomous evaluation agencies. Support could include capacity building activities that help countries build national assessment frameworks, and regional exchanges on building strong assessment institutions. Support could also include expanding the reach of tools to diagnose learning assessment systems and recommend improvements, and reviewing what works in public examinations reform and competency-based assessment initiatives, and support for innovation across groups of countries to co-develop assessments of skills beyond literacy and numeracy.

Global goods that build expertise in learning assessment throughout the education system, from teachers to ministerial leaders. Support could include formulating guidelines for building teacher capacity for classroom-based learning assessments and building the technical capacity of ministry officials (upskilling). A review of good practices for implementing both could also support this area. Low-cost innovations in building sustainable expertise among ministry officials and teachers would also be useful. This could include creating global public goods where one country's learning is designed to benefit the learning of other countries.

Global goods that support learning assessment systems for the most marginalized. This includes FCACs and the most marginalized learners within countries. Activities could include support for FCACs that do not conduct learning assessments to participate in cross-national assessments that meet GPE's quality standards, and guidelines for harmonizing donor priorities for assessment. This could include a review of the value for money of learning assessments, and a review of what works for FCACs' participation in national and cross-national assessments. Investments in this area could also support mechanisms for helping countries use existing data for global reporting and developing low-cost, technology-based solutions for collecting data in FCACs and among students with disabilities or learners experiencing displacement because of conflicts.

Global goods that support the systemic use of learning assessment data. This overlaps somewhat with the KIX data thematic area, but the DCPs and experts consulted felt strongly that it should also be considered. Regional and international exchanges on good practices in presenting and using learning assessment data could be useful to DCPs. A review of how learning assessment data (classroom-based, examinations, and large-scale) that influence policy in low- and middle-income countries would be useful. A systematic test-intervene-retest approach using learning assessments could shed light on how countries can incorporate these assessments in their education programs. Coming up with new ways of communicating and visualizing data could improve how learning assessment data are used.

To ensure KIX investments in these opportunities respond to the needs and landscape, several areas of investment are required that include:

- Building capacity through knowledge transfer, capacity development and learning exchange on issues where there is a sufficient evidence base.
- Building evidence and evaluation of what works on topics where there are some solutions, but where more synthesis is needed to develop a solid evidence base.
- Innovation in learning assessment systems on topics where new thinking and solutions are needed.

1. Introduction

The effective use of learning assessment data is fundamental to education systems. Without data on how students are learning, it is impossible to know whether an education system is succeeding in providing quality education. Learning assessment has been described as "the process of gathering and evaluating information on what students know, understand, and can do in order to make an informed decision about next steps in the educational process" [9] When analyzed together with contextual data, learning assessments can be a powerful tool for improving national education systems. A learning assessment system encompasses all of the policies, institutions, assessment tools and practices that contribute to gathering and using information on how students are learning in the education system.[10] The components of a learning assessment system vary by country, and potentially include assessments designed and conducted by teachers to inform their instruction (henceforth classroom-based assessments), national examinations taken by all students, and sample-based assessments used primarily for policymaking (henceforth large-scale assessments). A strong learning assessment system can help students, their parents, teachers and school administrators track progress, and be used for intervening in cases where a child's learning is offtrack.

Learning outcomes are central to the global education agenda set out in the Sustainable Development Goals (SDGs) and the Education 2030 Framework for Action, so the measurement of these outcomes is also a priority. SDG 4[11] lays out an ambitious set of targets that includes improving learning outcomes for all, from early childhood to adulthood. Of the 10 education targets in SDG 4, half relate to learning and skills.[12] This emphasis on learning in the SDGs was the result of a major movement by governments, civil society, teachers' organizations, and donors, including the Global Partnership for Education (GPE), demanding a higher bar for progress in the education sector, and a commitment to getting better data on learning.[13]

There is now a global mandate for strengthening learning assessment systems in relation to the SDGs. Using existing data, it is possible to estimate how many children are failing to meet minimum learning standards in mathematics and reading—six out of 10 children, or 617 million children worldwide.[14] But this is a very rough estimate because recent national data on learning outcomes are not available for many low- and middle-income countries, including GPE developing country partners (DCPs). African countries, least developed countries, landlocked developing countries, small island developing states, and countries affected by conflict have the lowest rates of participation in large-scale assessments.[15] While global

[9] Clarke 2012, 1.
[10] Clarke 2012.
[11] SDG 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.
[12] SDG targets 4.1, 4.2, 4.4, 4.6 and 4.7, and their corresponding indicators, are related to learning outcomes.
[13] High-Level Panel of Eminent Persons on the Post-2015 Development Agenda. *A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development.* New York: United Nations, 2013; Learning Metrics Task Force. *Toward Universal Learning: Recommendations from the Learning Metrics Task Force*. Montreal and Washington DC: UIS and Brookings, 2013; SDSN Thematic Group 4. *The Future of Our Children: Lifelong, Multi-Generational Learning For Sustainable Development.* New York: SDSN, 2014.
[14] UNESCO Institute for Statistics. "6 Out of 10 Children and Adolescents Are Not Learning a Minimum in Reading and Math." Montreal: UNESCO UIS, 2017a.
[15] UNESCO Institute for Statistics. "Investment Case for Expanding Coverage and Comparability for Global Indicator 4.1.1." Montreal: UNESCO UIS, 2017b.

estimates can call attention to the problem, they must be underpinned by robust and relevant national learning assessment systems that are useful to the governments administering them.

GPE and many other development institutions have made efforts to strengthen learning assessment systems, including helping countries improve their systems and producing global goods that transcend national boundaries, and that can support GPE DCPs and non-DCPs alike[16]. Global goods are "Institutions, mechanisms and outcomes that provide near universal benefits, reach across borders and extend across generations"[17]: that is, tools, products and approaches —including data, assessment tools, standards and research outcomes—that, once developed as the outcome of one particular intervention, can be adapted to create a tool or approach that is applicable, with appropriate customization, to other contexts.[18] This discussion paper describes the efforts being made to strengthen learning assessment systems, identifies the knowledge and innovation gaps, and suggests areas where GPE could invest through its new Knowledge and Innovation Exchange (KIX) initiative to strengthen these systems.

2. Paper development and consultation process

GPE commissioned a senior author to review GPE documents and sector plans related to strengthening learning assessment systems, conduct a desk review, consult with DCPs and experts, and write this paper. Working with GPE's senior education specialist on learning assessment systems and other members of the GPE Secretariat, the author drafted an initial discussion paper in May 2018. This included a summary of the challenges for building strong learning assessment systems, existing global goods, gaps in global goods that emerged from the desk review, and suggestions for areas in which GPE could invest through KIX. The draft was reviewed by the GPE Secretariat. A revised discussion paper was translated into French and both versions were sent to DCP representatives and international experts representing academia, civil society, foundations, donors, and multilateral institutions.

Consultations were held during an Africa regional workshop in Dar es Salaam in July 2018 convened by the Teaching and Learning Educators Network for Transformation of UNESCO Dakar, and an Asia-Pacific regional workshop in Penang in September convened by the Network on Education Quality Monitoring in the Asia-Pacific of UNESCO Bangkok. These workshops brought together some 65 government officials from 30 DCPs, typically the directors of learning assessments, curricula and teacher training.

The GPE Secretariat received in-person feedback from 30 DCPs, and written feedback from seven DCPs and 24 international experts. The results from these consultations are presented alongside the research evidence. On issues where there was agreement across stakeholders, or where the feedback was clarifying content in the draft paper, the feedback was incorporated directly into this paper. Where there was conflicting feedback, the different viewpoints are discussed and presented in the paper, along with a rationale for recommending a certain way forward if necessary.

[16] See section 4 for a list of the many existing global goods that help strengthen learning assessment systems.
[17] UNESCO. "Policy paper 34: Fulfilling our collective responsibility: Financing global public goods in education." Paris: UNESCO, 2018.
[18] The International Commission on Financing Global Education Opportunity. *The learning generation: Investing in education for a changing world.* Washington DC: Education Commission, 2016.

3. Challenges to building strong learning assessment systems

There are many reasons why strong learning assessment systems are not in place, especially in the most vulnerable countries. Four main challenges to these systems emerged from research and consultations with DCPs and international experts.

The first is the quality of the assessment tools and the lack of technical expertise for assessment design, administration and analysis. This was noted as a major barrier to strong learning assessment systems. Having good quality measurement tools (tests) is important. Gauging the quality of a measurement tool includes examining its measurement precision (reliability) and the degree to which an instrument measures what it is intended to measure (validity). For tools used across cultural contexts, cross-cultural validity is also important.

Another aspect of quality is the appropriateness for a given cultural context. For example, written assessments administered in schools might be an appropriate way to measure reading in countries where all children are in school and able to read. But in countries where this is not the case, oral assessments might be more appropriate. Gathering data through household surveys can supplement what is gathered through school administration. DCPs and international experts, however, reported that countries often commit to an assessment before having a dialogue on these types of issues.

Highly trained technical experts are needed to design, administer and analyze data from large-scale learning assessments and examinations so that these meet quality standards. These experts may not be available within a country or the region, so international experts are needed (and hired at a much higher cost than local professionals). Stakeholders consulted for this paper noted it is sometimes difficult to obtain the full tools or methodologies for large-scale assessments, so it is difficult to build upon work that has already been done. Given these requirements, learning assessments can be costly. Each administration of a large-scale assessments costs on average US$500,000 per country for data collection and technical assistance.[19] The cost of an early grade reading assessments varies widely, ranging US$150,000–U$300,000.[20]

For policy makers, researchers and teachers, capacity building and training on any type of assessment was reported by stakeholders consulted for this paper to be often informal, short-term and of varying quality—and sometimes with no formal recognition or certification.

Classroom-based assessments are often a low priority because the available tools are of poor quality and teachers lack training on how to develop or use them. Schools and teachers also lack the resources and options for making improvements on the basis of the information that classroom-based assessments provide. This ultimately leaves teachers, students and their families guessing whether learning is progressing on track. Many DCPs noted that very large class sizes of 60–80 students makes it difficult for teachers to know their students well enough to evaluate their learning in a consistent way.

The second challenge is aligning the various types of learning assessments used, and positioning these assessments within national education systems. These challenges stem from enrollment, not learning

[19] UIS 2017a.

[20] Wagner, Daniel, Martin Lockheed, Ina Mullis, Michael O. Martin, Anil Kanjee, Amber Gove, and Amy Dowd. "The Debate on Learning Assessments in Developing Countries." *Journal of Comparative and International Education*, *42*(3):509-545, 2012.

outcomes, often being the goal of education systems. So, the incentive for investing in measuring and improving of learning outcomes is often lacking.[21] The disconnect between the curriculum, pedagogy and assessment is also large, with the curriculum often being too advanced for most learners, teacher training being based on theory and not practice, and assessment policies being narrowly focused on two or three domains).[22]

A disconnect sometimes exists within a learning assessment system between the multiple assessments done in a country. One DCP expert consulted said that teachers using classroom assessments were surprised that students were scoring much lower than they expected on examinations. Balancing the need for assessments with the overarching goals of an education system is also a challenge. Too much assessment can take valuable class time away from learning, and too little can deprive teachers and policy makers of the information to do their jobs.

The strong focus on high-stakes examinations, which often emphasize rote learning in only a few domains, prevents teachers from being able to teach to the entire curriculum. In countries where there are limited spaces in secondary education, examinations can be a force for exclusion rather than for improving learning. Data from examinations can be an available and useful source of information on all students. But these data are rarely used to inform policy making and school-level decisions.

The many funders involved in an education system, including domestic agencies and international donors, have different priorities. This lack of alignment can contribute to weak learning assessment systems. Funding for assessments can be inconsistent, which results in an ad hoc approach dependent on when donors have funding available for carrying out an assessment. This was not, however, the opinion of all DCPs and international experts. Some said budgets sufficiently covered the costs of assessments, and that these costs were anyway small compared with overall education budgets. Lack of communication between the various departments involved in making decisions in ministries of education was also said to be a challenge.

The third challenge is that learning assessments often do not encompass the most marginalized children, including those with disabilities and those who are not in school. Even when a country has national assessment results, it still might not have sufficient data to analyze how different groups of students are learning—boys and girls, the wealthy and the poor, those with disabilities, those displaced by conflict, and children who do not speak the dominant language at home. To be useful, data from large-scale assessments and public examinations should be analyzed with contextual factors to determine how different groups of students are learning. Often data are collected but not sufficiently analyzed to provide insights on how to intervene.

Appropriate classroom-based assessments for children with disabilities are often lacking, and the resources to make the required modifications to the environment and pedagogy are often insufficient. As discussed earlier, large class sizes, which tend to be more common in marginalized communities, can prevent teachers from being able to assess learners and use the data to make improvements [23] Low

[21] Pritchett, Lant. “Creating Education Systems Coherent for Learning Outcomes: Making the Transition.

[22] UNESCO International Institute for Capacity Building in Africa. *Teaching policies and learning outcomes in Sub-Saharan Africa: Issues and Options.* Addis Ababa: UNESCO IICBA, 2016; Anderson, Kate, Seamus Hegarty, Martin Henry, Helyn Kim, and Esther Care. *Breadth of Learning Opportunities: A Fresh Approach to Evaluating Education Systems.* Washington DC: Brookings Institution, 2018.

[23] UNESCO. “Global Education Monitoring Report 2010: Reaching the Marginalized.” Paris: UNESCO, 2010.

attendance rates of both teachers and students can also prevent teachers from collecting regular classroom-based assessments.

The fourth challenge is the lack of use of learning assessment results. This is influenced by how assessments are designed and communicated, and the expertise and resources available for making change. An important attribute of a good assessment is that it is "fit-for-purpose," meaning it provides the needed information to those using it.[24] The purposes of learning assessments include supporting learning in the classroom, system-level policy making, accountability, evaluating special projects and pilots, and for certification, progress and transfer. Ideally, these purposes are determined before starting an assessment, and education systems devote sufficient resources across the different components of a learning assessment system. In practice, however, the emphasis is typically on assessments with the highest stakes, usually end-of-cycle examinations.[25] Some components of a learning assessment system can also serve more than one function, so not all of them require costs being incurred again for each new type. The multiple uses of assessments, however, are not typically leveraged, according to stakeholders consulted for this paper.

A concern mentioned by nearly every DCP and international expert was that learning assessment results are not used in ways that support improvements in learning. They reported a lack of connection (feedback loop) between data and practice. A similar concern was cited for classroom-based assessments, because the resources and expertise to make changes were limited. At the system-level, this can also because both are lacking. It is also possible that no staff are available to analyze the data, or the data are analyzed but not getting into the hands of those who can make the changes. A further reason is that MOEs sometimes collect data first and later decide how to use it, rather than starting with the questions they want to answer.

Another contributing factor is that data on learning are often analyzed in terms of mean scores, with the goal of raising national averages or getting more students into the "proficient" category. This excludes struggling and advanced learners.[26]

4. How GPE supports learning assessment systems

Because of these challenges, GPE has made building strong learning assessment systems a priority, as shown by its results framework and funding model. Improved and more equitable learning outcomes is first of the GPE 2020 Strategic Plan's three goals,[27] and the GPE results framework and funding model thus require monitoring of learning outcomes. In 2015, GPE developed its first results framework, for 2016–2020.[28] The framework has one indicator for learning outcomes and two for learning assessment systems. These are:

[24] Archer, Elizabeth. "The Assessment Purpose Triangle: Balancing the Purposes of Educational Assessment." Frontiers in Education. 2 (41), 2017.
[25] Wagner et al. 2012; Learning Metrics Task Force. *Toward Universal Learning: Implementing Assessment to Improve Learning.* Montreal and Washington DC: UIS and Brookings, 2014.
[26] Wagner, Daniel, Sharon Wolf and Robert F. Boruch, eds. *Learning at the Bottom of the Pyramid: Science, Measurement and Policy in Low-Income Countries.* Paris: UNESCO IIEP, 2018.
[27] Global Partnership for Education. 2016. "GPE 2020 Strategic Plan." Global Partnership for Education. Washington DC: GPE, 2016.
[28] Global Partnership for Education. "GPE Results Framework for 2016-2020." Global Partnership for Education. Washington DC: GPE, 2017.

Indicator 1: proportion of DCPs showing improvement on learning outcomes (basic education). Using data from nationally representative school-based learning assessments, this indicator shows how many countries have had a statistically significant increase in primary and lower-secondary learning outcomes.

Indicator 15: proportion of DCPs with a learning assessment system within the basic education cycle that meets quality standards. These standards include the enabling context in which the assessment is conducted, the quality of assessment tools and processes, and the degree to which the assessment is aligned with the national curriculum.

Indicator 20: proportion of grants supporting education management and information systems and/or learning assessment systems. This indicator refers to the number of education sector program implementation grants (ESPIGs) that support learning assessment systems and/or education sector management and information systems.

Annex A gives more information on these indicators, and the progress made by DCPs.

Using data collected for indicator 20, it is possible to determine the number of DCPs receiving GPE support to strengthen learning assessment systems. As of June 2017, 29 (or 71 percent) of the 41 active ESPIGs were investing in learning assessment activities.[29] Activities supported by these grants include the development and implementation of classroom-based assessment and national assessment in Bangladesh, and setting up an independent agency in charge of national assessments in the Democratic Republic of Congo.

The GPE funding model (adopted in 2014 by the Board for the 2015–2018 replenishment period) requires countries applying for an program implementation grant to either have a system in place to monitor learning outcomes or a plan to develop one.[30] In 2017, the Board decided that if funding is insufficient to build these systems, GPE funding should be used. The new funding model also allows GPE to provide results-based financing, which incentivizes countries to set and achieve their own learning targets. To receive the first 70 percent of GPE funding, DCPs must meet several requirements, as explained earlier. The disbursement of the remaining 30 percent is linked to demonstrated progress on sector results, including learning outcomes. Governments, in consultation with their partners in the local education groups, must identify a transformational strategy to improve learning outcomes that outlines actions to remedy issues driving low learning levels. For example, the Democratic Republic of Congo has linked funding to improved reading performance in primary education. A commitment to increase learning outcomes is therefore one way that DCPs can access the partnership's results-based financing.

GPE has also supported global and regional activities to strengthen learning assessment. The global and regional activities program, which has formally concluded, contained two grants focused on learning assessment systems, totaling nearly US$1.5 million. The first grant supported a UIS project from 2013 to 2015 to develop methodologies to link reading assessments across regions, identify best practices for early reading assessment and initiate a global catalogue of learning assessments. The second grant supported the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP) to build regional evidence and capacity. This was carried out by UNESCO Bangkok from 2014 to 2016.

[29] This number does not include sector pooled grants.
[30] Public examinations and issuing diplomas do not count toward this requirement.

The A4L initiative was launched in July 2017 to build capacity for national learning assessment systems to measure and improve learning as a pilot for the Knowledge and Good Practice Exchange approach. A4L has a budget of about US$3 million and three components: (1) tools to support diagnostics of learning assessment systems, to be made publicly available after piloting in three DCPs in 2018 and 2019; (2) support to NEQMAP and the Teaching and Learning Educators' Network for Transformation in Sub-Saharan Africa for capacity development, analytical work and knowledge sharing; and (3) a landscape review on measurement of 21st century skills and tools to support such measurement (with the Center for Universal Education at Brookings).

GPE is strongly supporting learning assessment systems by incorporating them in the GPE 2020 Strategic Plan, supporting DCPs to monitor learning outcomes, supporting global and regional efforts to improve learning assessments, and building capacity through the A4L initiative. A4L in particular has set the stage for KIX by supporting a system approach to strengthening learning assessment systems. GPE is developing a tool for countries to analyze their learning assessment systems and thereby feed into efforts to strengthen them. The partnership is also supporting regional networks to facilitate expertise for countries to build learning assessments that respond to national curricula and policy priorities. KIX grants will build on these efforts by further enabling knowledge exchange, supporting new global goods and building on the existing ones, which are discussed in the following section.

5. Global goods and innovation in learning assessment

Global goods for learning assessment in developing countries fall in to three general categories: large-scale assessments of learning, capacity building tools and initiatives, and innovations in the use of learning assessments.

Large-scale assessments of learning

The global goods in this category include tools and studies, data and networks, and knowledge sharing mechanisms.

Tools and studies

Large-scale assessments are used for monitoring and providing information to policy makers and practitioners on overall performance levels, how these levels change, and how various characteristics of the learner, family, school and community are related to performance levels. These assessments can have multiple aims, but they typically provide information on the overall health of the system, and are used for making decisions on policies and resource allocation.

Large-scale assessments can be national or cross-national. National assessments are specific to each country and are typically aligned with the national curriculum. UNESCO and UNICEF's Monitoring Learning Achievement (MLA) initiative was one of the first global goods supporting national learning assessment systems in low- and middle-income countries. In a follow-up to the 1990 World Conference on Education for All in Jomtien, Thailand, UNESCO and UNICEF piloted an approach to develop national assessments with China, Jordan, Mali, Mauritius and Morocco in the mid-1990s.[31] The resulting guidelines and approach was used throughout sub-Saharan Africa and Asia, and several DCPs continue to use revised

[31] Chinapah, Vinayagum. *Monitoring Learning Achievement: Toward Capacity Building.* Paris: UNESCO, 1997.

versions of the MLA approach as their national assessment tools.[32] Cross-national assessments allow for comparison between participating countries. They are often developed through a consensus process, taking into account the national curricula of the participating countries. Both national and cross-national assessments can also allow for comparison over time if designed for that purpose.

If a country is interested in measuring learning by adapting an existing tool or participating in a study with other countries, there are many tools that can be considered global goods, as Table 1 shows. These tools include internationally and regionally comparable studies. Assessment approaches that are publicly available for adaptation include early grade reading and mathematics assessments, (EGRA/EGMA) and citizen-led assessments (CLAs). These do not provide comparable data across countries. But they do enable economies of scale for countries interested in assessing learning because they do not need to build an assessment from scratch. Another relevant effort is the foundational learning skills module from UNICEF's Multiple Indicator Cluster Survey (MICS), a new household survey designed to provide data for SDG 4.1.1(a). The results are globally comparable and will be available for a substantial number of DCPs in late 2018. This assessment reaches out-of-school children, enables a detailed equity analysis based on the extensive sociodemographic information from households, and allows for cross-sectoral analysis because of the information on nutrition, health, child labor, and child discipline.

Table 1: Multicountry Learning Assessments

	Name of Assessment	Acronym	Grade or age assessed	Subject or domain assessed	Administering agency
Results can be compared cross-nationally	Progress in International Reading Literacy Study	PIRLS	Grade 4	Reading	International Association for the Evaluation of Educational Achievement (IEA)
	Trends in International Mathematics and Science Study	TIMSS	Grade 4 Grade 8	Mathematics Science	IEA
	International Civic and Citizenship Study	ICCS	Grade 8	Civics and citizenship, Global citizenship	IEA
	Programme for International Student Assessment	PISA	15 year-olds	Reading Mathematics Science optional domains (e.g. collaborative problem solving in 2015)	Organisation for Economic Co-operation and Development

[32] Friedman, Tim, Ursula Schwantner, Jeaniene Spink, Naoko Tabata, and Charlotte Waters. *Improving Quality Education and Children's Learning Outcomes and Effective Practices in the Eastern and Southern Africa Region.* Nairobi: UNICEF ESARO, 2016.

Results can be compared cross-nationally	Latin American Laboratory for Assessment of the Quality of Education	LLECE	Grade 3 Grade 6	Reading and writing Mathematics Natural sciences (Grade 6)	UNESCO-Regional Bureau for Education in Latin America and the Caribbean
	Multiple indicator cluster survey foundational learning skills module*	MICS	7–14 year-olds	Reading Number skills	UNICEF
	Programme d'analyse des systèmes éducatifs de la CONFEMEN	PASEC	Grade 2 Grade 6	Reading Mathematics	CONFEMEN
	Pacific Islands Literacy and Numeracy Assessment	PILNA	Grade 4 Grade 6	Literacy Numeracy	Education Quality and Assessment Program
	Southern and Eastern Africa Consortium for Monitoring Educational Quality	SEACMEQ	Grade 6	Reading Mathematics HIV/AIDS knowledge	SEACMEQ
	Literacy and numeracy assessment	LaNA	Grade 4 (or 5 or 6 if appropriate)	Basic reading Basic mathematics	IEA
	Programme for International Student Assessment for Development*	PISA-D	15 year-olds (includes out-of-school children in most countries)	Reading Mathematics Science	PISA
	Southeast Asia Primary Learning Metric*	SEA-PLM	Grade 5	Reading and writing Mathematics Global Citizenship	Southeast Asian Ministers of Education Association/ UNICEF East Asia Pacific Regional Office

Results cannot be compared cross-nationally	Citizen-led assessments (Annual Status of Education Reports, Uwezo, etc.)	CLA	Varies by country, typically 5–6 to 16–18 year-olds	Basic reading Basic mathematics	Varies
	Early grade reading assessment	EGRA	Grades 1–3	Basic literacy	Varies
	Early grade mathematics assessment	EGMA	Grades 1–3	Basic mathematics	Varies

* In field testing stage; cross-national comparisons may be established in the future.

While the learning domains most frequently assessed worldwide in national and cross-national assessments are literacy and numeracy, a strong learning assessment system would include most of the skills and competencies articulated in the national curriculum. In most countries this is much broader than these two domains.[33] The Southeast Asia Primary Learning Metric (SEA-PLM), Programme for International Student Assessment (PISA), and the International Association for the Evaluation of Educational Achievement (IEA)'s International Civic and Citizenship Study, among other studies, include domains related to global citizenship and collaborative problem solving. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SEACMEQ) includes HIV/AIDS awareness and life skills.

More than 80 percent of countries worldwide conducted some type of large-scale assessment of reading and mathematics between 2010 and 2015.[34] Of the 65 GPE DCPs during this period, 50 (77 percent) participated in some type of national, regional or international assessment. By including EGRA, EGMA and citizen-led assessments, the number of participating DCPs in any of these types of assessments rises to 55 (85 percent).[35] Annex B lists the DCPs and the assessments in which they participate.

Early grade reading and math assessments are the most common types of assessments in which DCPs participate (41 countries or 63 percent), although these are not always done on a nationally-representative sample. Even so, they provide a set of tools on which a larger study could be built. National assessments are also commonly done in DCPs (39 countries or 60 percent), as well as regional assessments (25 countries or 38 percent). Citizen-led assessments and international assessments are conducted in 11 DCPs (14 percent for each type). Nine DCPs have not participated in any of the listed assessment types since 2010: Central African Republic, Comoros, Djibouti, Dominica, Grenada, Mauritania, Saint Lucia, Saint Vincent and the Grenadines, and Uzbekistan.[36] Many of these are small island developing states or FCACs, which may lack the infrastructure and resources to develop or participate in large-scale learning assessments.

[33] Care, Esther, Helyn Kim, Kate Anderson, and Emily Gustafsson-Wright. *Skills for a Changing World: National Perspectives and the Global Movement.* Washington DC: Brookings Institution, 2017.
[34] UNESCO Institute for Statistics 2018.
[35] EGRA and EGMA studies are often used for program evaluation and therefore not always done on a nationally representative sample.
[36] Uzbekistan plans to participate in PISA in 2021.

Data

The UIS.Stat website is the main source of data for the education sector. It includes available data on all SDG 4 indicators. All of the assessments listed in the previous section offer participating countries the opportunity to publish their results data through a central online platform. This is sometimes in the form of individual country reports or syntheses of findings across countries and interactive reports. Some online platforms also include data sets. For example, USAID launched in 2018 the Early Grade Reading Barometer, which provides reports and data on countries that have used versions of the EGRA tools.[37] This website displays data in multiple ways, including an overview of how well students are reading in a country, percentage of struggling readers, student performance on EGRA subtasks, and comparison of EGRA results across regions. The People's Action for Learning Network offers downloadable datasets from many of the countries participating in the citizen-led assessments.

A source of metadata on learning assessments is the UIS Catalogue of Learning Assessments database, which was developed with GPE funding. This has information on various aspects of nationally representative learning assessments in countries around the world. The catalogue includes assessments conducted in schools and in homes. It covers assessments conducted in early childhood education, primary, and secondary schooling, and the corresponding age groups to these educational levels (for household-based assessments). The Learning Assessment Capacity Index shows these results in a global map. Using this and several other data sources, it is possible to identify the types of assessments that DCPs engage in (Annex B).

Networks and knowledge sharing

There are several forums for those working on large-scale assessments to network and share knowledge. Each of the large-scale assessments listed in Table 2 has some type of networking mechanism, usually involving regular in-person meetings and virtual forms of collaboration. IEA has organized most of the data and documents from all cross-national assessments in an online platform, the International Large-Scale Assessment (ILSA) Gateway.

The Global Alliance to Monitor Learning (GAML) is an institutional platform set up by the UIS to harmonize standards and coordinate efforts globally for measuring learning. GAML brings experts together to develop methodologies on learning indicators to track progress on SDG 4 and to set standards for good practices on learning assessments. While GAML's products are global goods, it is also the only network that brings together nearly all stakeholders in learning assessment globally. The networking that takes place through being a GAML member can also be considered a public good. The International Institute for Education Planning (IIEP) learning portal is a global platform that provides a knowledge bank of resources, blogs, and webinars on learning and educational planning, including many documents on monitoring learning.

The International Association for Educational Assessment (IAEA) and the National Council on Measurement in Education (NCME), among other professional organizations, convene researchers and government officials for knowledge sharing and cover a range of topics in assessment. At the regional level, the NEQMAP and TALENT efforts (in the Asia-Pacific and Sub-Saharan Africa respectively) facilitate capacity development, knowledge-sharing and produce cross-country research studies. A similar effort is the Network for African Learning Assessment (NALA). This was initiated by the Association for the Development of Education in Africa (ADEA) and the governments of Kenya, Rwanda, Senegal and Zambia

[37] USAID. "Early Grade Reading Barometer." Accessed May 8, 2018. http://www.earlygradereadingbarometer.org/.

to facilitate peer learning on assessment.[38] Regional professional organizations such as the Association for Educational Assessment in Africa (AEAA) also provide platforms for regional knowledge sharing, and typically among examinations bodies.

Capacity development

A second category of global goods covers efforts to build capacity for learning assessment. This category includes tools and publications that support countries in measuring learning, information on system capacity and knowledge sharing.

Tools and publications

The five-volume National Assessments of Educational Achievement series, published by the World Bank from 2008 and 2015 includes guidance on building a national assessment system, developing tests and questionnaires, implementing assessment systems, using assessment information, and analyzing data. The publication is the most comprehensive resource on for nationally assessing education achievement. It is available online and the World Bank also provides free print versions, some of which come with CD-ROMs with additional resources.[39] The publication has been translated into in Arabic, French, Portuguese, Russian, and Spanish. The World Bank also offers an open access e-learning course on student learning assessment. This step-by-step introductory course was developed specifically for policy makers and practitioners in developing countries.[40] UNESCO IIEP has a 10-volume series on quantitative research methods in educational planning, which covers research and sample design, item writing, trial testing, and item analysis, among other topics. It also hosts a Massive Open Online Course (MOOC) on learning assessments.

To improve the relevance and cohesion of the assessment system, GAML developed the so-called principles of good practice in learning assessment. This describes how to develop and implement large-scale assessment programs so that the data can be used for education system monitoring and evidence-based education policy. The good practice in learning assessment document provides principles relevant to the various large-scale learning assessment activities being undertaken throughout the world, including cross-national and national assessments. GAML has also begun publishing a series of Quick Guides with step-by-step instructions for countries to assist them to report on the SDG 4 indicators.[41]

An example of an effort for national capacity building that also produced global goods is the World Bank's READ Trust Fund. Since 2008, READ has provided about US$38 million in support to country assessment systems. The lessons from these experiences have been used to develop a series of global knowledge products on assessment.

Information on system capacity

The World Bank's SABER Student Assessment program offers resources and data for gauging the overall quality of an educational assessment system. The diagnostic toolkit consists of standardized questionnaires and rubrics that allow users to map the characteristics of learning assessments with quality standards for all types of assessment activities done by a government. This includes national

[38] Mugo, John. "Welcoming Two New Initiatives to Accelerate Quality Teaching and Learning in Africa." Washington DC: Brookings Institution, 2017

[39] Greaney, Vincent, and Thomas Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Washington, DC: World Bank, 2008.

[40] The series is at https://openknowledge.worldbank.org/handle/10986/2143.

[41] Ramirez, Maria Jose. *Quick Guide No. 3 Implementing a National Learning Assessment.* Montreal: UIS, 2017.

examinations, national sample-based assessments, international and regional assessments, and classroom assessments. The UIS Learning Assessment Capacity Index (LACI) has a map showing whether countries are demonstrating various levels of capacity for large-scale assessments on the basis of their participation in national, regional and international assessments.[42]

GPE is developing a complementary tool to support diagnostics of learning assessment systems called the Analysis of National Learning Assessment Systems (ANLAS). This is specifically aimed at informing strategies to enhance these systems, and to be operationalized through countries' education sector plans. The ANLAS toolkit is the first diagnostic tool that will include analysis of the extent to which learning assessment systems consider 21st century skills. The tool will be piloted in three DCPs in 2019 and then be finalized and disseminated.

Networks and knowledge sharing

The international and regional assessments include capacity building as a component of the participation in the study.[43] IEA literacy and numeracy assessment for developing countries was created to build capacity for learning assessment at the local level and support the efforts of the global education community.[44] The Organisation for Economic Co-operation and Development (OECD) also has resources for building capacity in developing countries. Countries participating in PISA-D and some PISA participants have conducted capacity-needs analysis based on SABER Student Assessment rubrics and OECD quality standards. These countries have also worked with the OECD to develop a capacity building plan to support their successful participation in the PISA-D assessment.[45] To date, PISA-D capacity needs analyses and/or capacity building plans have been completed for Cambodia, Ecuador, Guatemala, Honduras, Panama, Paraguay, Senegal, Tanzania, Ukraine and Zambia. While these efforts are promising, they are nevertheless tied to specific studies. Increased networking and knowledge sharing is needed to develop capacity independent of a country's decision to participate in international assessments.

Innovations in learning assessment

The third category of global goods are initiatives breaking new ground in learning assessments.

Tools

The UIS is exploring approaches to equate learning outcomes across countries that use different learning assessments for SDG 4 reporting. An initial stage in this process was to develop common reporting scales in collaboration with the Australian Council for Educational Research (ACER). These numerical scales are associated with substantive descriptions that explain levels of proficiency in the learning domains identified in the SDG indicator framework. For example, for indicator 4.1.1 on minimum proficiency levels in reading and mathematics, the substantive descriptions provide an increasingly complex set of skills for reading and mathematics. Particular locations on the UIS reporting scales would be established as benchmarks to enable common global definitions of terms, such as "minimum proficiency" and "developmentally on track." The Learning Progression Explorer is an online tool that will allow users to explore these reporting scales. This innovation represents a new way of equating learning outcomes across different countries and the educational trajectory of learners (that is, from early childhood through secondary education).

---

[42] UNESCO Institute for Statistics 2018.

[43] Lockheed, Marlaine E. *The Craft of Education Assessment: Does Participating in International and Regional Assessments Build Assessment Capacity in Developing Countries? An Independent Evaluation of IEA's Program on the Assessment of Student Achievement (PASA),* 2010.

[44] IEA. "IEA Studies." Accessed May 08, 2018.

[45] OECD. *Building Capacity for Assessment in PISA for Development Countries.* Paris: OECD, 2017.

The process of developing the reporting scales has sparked much debate on whether it is possible to have a truly universal scale across multiple countries, languages and contexts. Therefore, GAML is also exploring a "social moderation" approach developed by Management Systems International. Here, groups of national experts, including teachers, evaluate national and cross-national assessments to determine how they equate to descriptions of proficiency levels.

Research

Few global goods exist for classroom-based assessment because classroom practice and classroom-based assessment are often seen as the responsibility of teachers and school leaders, and are highly dependent on the curriculum and instructional needs of national education systems. There are, however, examples of simple learning assessments being used to gather information on learners and plan interventions that have resulted in improved learning. For example, the Read India Program, implemented by the nongovernment organization Pratham, used the Annual Status of Education Report's literacy and numeracy assessments to group children by ability level. Teaching to their individual learning levels resulted in sustainably improved outcomes for some children.[46] This program has expanded in recent years by providing tablets with reading and math games, and adding a parental involvement component.[47]

There are numerous examples of how countries use data from large-scale assessments to inform policies. Best et al. (2013) found that the results of large-scale assessments in developing countries were most frequently used for education policy making in the areas of resource allocation, curriculum standards and reform, performance, standards, and assessment policies. Teaching and learning practices, community and parent engagement policies, and accountability policies were the policy areas with the least indication of use. Barriers to the use of assessment data found most commonly across studies included the quality of the assessment program and lack of in-depth analyses of the data. In Asia and the Pacific, assessments are most frequently used by policy makers for monitoring and evaluating education policies, followed by informing policy implementation, agenda-setting and new policy formation.[48]

The Optimizing Assessment Project, initiated by Brookings in 2017 in partnership with GPE, is an example of an initiative that draws on global good practices to build capacity for classroom assessments. The project is focused on Africa and Asia and works with regional networks and governments to develop classroom assessments for 21st century skills.[49] The Breadth of Learning Opportunities initiative of Education International and Brookings provides tools to measure the alignment in learning opportunities from the perspective of the teachers and school administrators, and at policy levels.[50] These open-source tools include questions about classroom assessments, national examinations and large-scale assessments in which a country participates.

[46] Banerji and Walton 2011.
[47] Winthrop, Rebecca, Eileen McGivney, and Adam Barton. *Can We Leapfrog? The Potential of Education Innovations to Rapidly Accelerate Progress.* Washington DC: Brookings Institution, 2017.
[48] Tobin, Mollie, Petra Lietz, Dita Nugroho, Ramya Vivekanandan, and Tserennadmid Nyamkhuu. *Using Large-scale Assessments of Students Learning to Inform Education Policy: Insights from the Asia-Pacific Region*. Melbourne and Bangkok: ACER and UNESCO Bangkok, 2015.
[49] Care, Vista and Kim 2018.
[50] Anderson et al. 2018.

6. Gaps in available global goods

Most existing global goods focus on (1) global goods that support groups of countries to design and administer specific large-scale learning assessments, (2) global goods that bring together networks of professionals working on assessment issues, and (3) online resources with general guidelines on how to develop quality learning assessments. Stakeholders participating in the consultations for this paper generally said that section 5 contained a comprehensive summary of the available global goods for strengthening learning assessment systems.

The perceived gaps in existing global goods varied by constituency. DCPs overwhelmingly said there was a gap in guidance for training teachers to conduct and use classroom-based assessments, and in good practices in examination reform. They also mentioned a lack of guidelines on how to package evidence from learning assessment in a way that is useful for policy makers, teachers and other stakeholders. While participation in large-scale assessments provides some capacity building for some ministerial officials, they noted that this needs to be more widespread to build an overall culture of evaluation. DCPs also noted inconsistencies across their learning assessment systems, and requested guidance on how to build strong institutions to oversee the entire national assessment portfolio.

The international experts were more concerned with a lack of globally comparable data on learning outcomes, and were interested in getting more countries to participate in cross-national assessments. Many noted additional budgetary support is needed for some regional assessment initiatives.

7. Potential investment areas

Through the research and consultation process, four main opportunities emerged for which global goods funded by KIX could be critical. These opportunities also align with GPE's results framework and funding model. They are:

Opportunity 1: Global goods that support national learning assessment institutions.

These include MOEs and autonomous and semi-autonomous evaluation agencies. One area of need is capacity building activities that help countries build national assessment frameworks. Consultation participants confirmed a strong assessment framework is an approach where (1) learning and learning outcomes are improved through the effective use of classroom assessments that inform pedagogical practices; (2) examinations, if used, certify the attainment of this learning; and (3) large-scale assessments are used to monitor and support systems and processes to improve learning and teaching within the system.[51]

One way to build capacity in this area is through regional exchanges on building strong assessment institutions. Many regional efforts for sharing learning that could host such an exchange are already underway. These could also provide an opportunity to expand the reach of tools to diagnose learning assessment systems and be used for recommending improvements, such as the SABER Student Assessment and GPE ANLAS.

[51] Kanjee, Anil. Excerpt from consultation feedback. 10 August 2018.

A major opportunity for KIX is guidance on reforming national examination systems. In many countries, examinations are of poor quality and constrain efforts in other areas of assessment.[52] In some education systems, limited spaces in secondary and tertiary education cause public examinations to be a tool used to exclude students from higher levels of education. Public examinations also often set the bar too low for what learners should achieve by the end of their schooling. For example, a 2017 study of public examination papers in Uganda, Nigeria, Pakistan and India found that most examination questions covered lower-order skills, such as recalling facts, and only a few address higher-order skills, such as reasoning.[53] There is correlational evidence that countries with upper-secondary examinations have higher student performance levels on international assessments, such as TIMSS and PISA. But there is a need to further unpack the links between examinations and learning levels, especially in developing countries. [54]

National examinations tend to receive the most domestic resources for educational assessment, as they are done annually. They are intended to be administered to an entire population of students and often have government or quasi-governmental organizations dedicated to exam development, administration and certification. The most common complaint about national examination systems in many developing countries is that they drive teacher practices, and lead to a narrowing of the curriculum to only the subjects covered by the exam.

Examinations are intended to certify successful completion of schooling and are an indication of the government's commitment to provide quality education to all citizens. They also enable access to higher education, employment and a way out of poverty. Because of their importance, cheating, fraud and malpractice are widespread in the examination systems of some DCPs.[55] Falsifying certificates is common, and the examination industry is rife with bribery and corruption.[56] Students often feel that examination results do not accurately reflect their knowledge and capabilities. And employers are increasingly bypassing the education examination system in some sectors; for example, information technology, where credentialing information is available from third-party online platforms.[57] That said, eliminating or even reforming national examinations is politically unfeasible in most countries as many benefit from the system. In some very corrupt countries, examination systems are, though corrupt, among the least corrupt institutions, and governments are reluctant to let go of these "islands" of relative non-corruption. Compiling evidence on alternatives to the outdated examination systems is an area where global goods are needed.

There are also domains identified as important for work and life, but under-assessed in education systems. While learning assessments tend to focus on literacy and numeracy, the curricula in most countries is substantially broader.[58] Important yet under-assessed areas include 21st century skills, social and emotional learning, and global citizenship. Work is ongoing to develop tools for assessing learning in these

[52] Learning Metrics Task Force 2014; Burdett, Newman. "Review of High Stakes Examination Instruments in Primary and Secondary School in Developing Countries." RISE Working Paper 17/018, 2017.
[53] Burdett 2017.
[54] Clarke 2012.
[55] Petters, Janet Sunday, and Maureen O. Okon. "Students' Perception of Causes and Effects of Examination Malpractice in the Nigerian Educational System: The Way Forward for Quality Education." *Procedia - Social and Behavioral Sciences* 114: 125-29, 2014.
[56] Transparency International. *Global Corruption Report: Education*. London: Earthscan, Publishing for a Sustainable Future, Routledge, Taylor & Francis Group, 2013; Mutambo, Aggrey. "UK Company Fined Sh330 Million for Bribing Kenya Officials." *Daily Nation*. January 11, 2016.
[57] Winthrop, McGivney and Barton 2017.
[58] Care et al. 2017.

areas, though some areas are more advanced than others—and some areas are inherently more difficult. The predictive value and helpfulness of large-scale assessments in some areas are being debated, including 21st century skills, such as critical thinking, collaboration, problem solving and creativity.

On a related topic, policy shifts are underway in many countries toward competency-based curricula and learning assessment. The term competency-based assessments is widely used, yet its meaning is open to various interpretation. A review of good practices in competency-based assessment could be a useful global good for many DCPs implementing a competency-based curriculum.

Opportunity 2: Global goods that build expertise in learning assessment throughout the education system, from teachers to senior government officials.

The issue here is not a lack of know-how, but rather the lack of political will and funding to scale up what has been shown to work. A useful global good would be guidelines for building the technical capacity of ministry officials (upskilling) that are sustainable and include real-time technical assistance when they need it.

Low-cost innovations in building sustainable expertise among ministry officials and teachers could also be useful. These could include creating global public goods where one country's learning is designed, from the beginning, to benefit the learning of other countries.

Building evidence is needed for increasing teacher capacity for classroom-based assessment. This includes assessments done by teachers and learner self-assessments, such as self-evaluations and portfolios. When done frequently (at least weekly), classroom-based assessments can be effective for improving learning outcomes.[59] DCPs seem interested in building capacity among school leaders and teachers for using continuous assessment to improve learning. For example, the Learning Champions initiative led by Brookings found strong interest in continuous assessment from government and civil society in Ethiopia, Pakistan, Palestine, Senegal, Tunisia and Zambia—and a lack of resources for officials trying to support these assessments in this area .[60] It is necessary to expand the knowledge base on how teachers can best be prepared to use classroom-based assessment to improve learning, specifically in DCP contexts.

Another area in which expertise could be built is in enabling education systems to develop new assessments in all of the curricular areas they want to assess. Innovations in this area would ideally encompass both traditional academic skills and under-assessed domains, such as (but not limited to) critical thinking, problem solving and creativity.

Opportunity 3: Global goods that support learning assessment systems for the most marginalized.

Learning assessment systems are lacking in the most marginalized countries, especially FCACs, and for the most marginalized learners within countries. In the most vulnerable countries, there is often a lack of nationally representative data to support interventions that improve learning.[61] Global goods are needed to support FCACs that do not conduct assessments to participate in cross-national assessments that meet GPE's quality standards. These global goods could also include reviews of the value for money of assessments and what works for FCACs' participation in national and cross-national learning assessments.

[59] Black, Paul, and Dylan Wiliam. "Developing a Theory of Formative Assessment." In J. Gardner, John, ed. Assessment and Learning. London: SAGE, 2006.
[60] Anderson, Kate, and Joshua Muskin. *Learning Champions: How 15 Countries, Cities, and Provinces Came Together to Rethink Learning Assessment.* Washington DC: Brookings Institution, 2018.
[61] UIS 2017a.

Many choices are available to countries on learning assessments, but little is known about the value for money of the different assessment approaches. For countries that have no assessment system in place, or countries where learning and assessment have been interrupted by conflict or fragility, it is important to know which types of assessment yields the most relevant information for the lowest cost. Research is needed on how best to choose assessment options where there are constraints to funding, human resources and capacity.

The current global goods for learning assessment generally exclude specific populations, such as children with disabilities, and out of school and displaced populations. Innovations are needed to capture learning for all children. Many stakeholders consulted for this paper mentioned the difficulties faced by FCACs in streamlining donor priorities for assessment so that the system works for their needs. Guidelines for harmonizing donor priorities for assessment could be a useful global good.

Progress is being made by UIS and the Australian Council for Educational Research on equating international and regional assessments. But there is still no internationally comparable baseline on learning outcomes to track progress on SDG 4. This restricts opportunities for cross-national comparison and peer learning, especially in developing countries. Establishing a global baseline will require a collective effort by national governments, regional assessment initiatives, and international organizations. Further innovation is needed, especially on getting data for as many countries as possible without overburdening them or causing scarce resources to be used for assessments that end up not being used. Investments for this could also support mechanisms for helping countries use existing data for global reporting and developing low-cost, technology-based solutions for collecting data in FCACs and among students with disabilities or learners experiencing displacement.

Opportunity 4: Global goods that support systemic use of learning assessment data.

This has some overlaps with the KIX data thematic area, but DCPs and experts consulted felt strongly that it should also be considered. Regional and international exchanges on good practices in presenting and using learning assessment data could be useful to DCPs. A review of how learning assessment data (classroom-based, examinations, and large-scale) influence policy and practice on teaching and learning in low- and middle-income countries would also be a useful contribution.

Some countries have developed ways of deriving information from examinations for improving classroom instruction. Kenya is one of them.[62] Because examinations receive considerable assessment-related resources, much more work needs to be done on strengthening these and other assessments focused on certification, progress and transfer, and supporting countries to make better use of these mechanisms.

Large-scale assessments can be used to improve learning in the classroom, but this is often not done. In a review of how the results of large-scale national, regional and international assessments are used in 73 developing countries, only a few of them used the results to inform teaching and learning practices, such as learning strategies and student-oriented pedagogy.[63] Leveraging large-scale assessments to improve learning in the classroom is another area in which global goods could bring together the existing research to build an evidence base that would be useful for DCP policy makers and practitioners.

[62] Somerset, H.C.A. *Examination Reform in Kenya*. Washington DC: World Bank, 1987.

[63] Best, Maura, Pat Knight, Petra Leitz, Craig Lockwood, Dita Nugroho, and Mollie Tobin. "The Impact of National and International Assessment Programmes on Education Policy, Particularly Policies regarding Resource Allocation and Teaching and Learning Practices in Developing Countries." Melbourne: ACER, 2013.

Large-scale assessments are often not well-used because national assessments in particular can be of poor quality and do not always support comparisons over multiple years. There is sometimes a push by governments and donors to develop and pilot new assessment tools, albeit without examining ways to "reuse" the results from earlier assessments. More evidence should be gathered on how to conduct different analyses and create innovative approaches to using learning assessment data for multiple purposes.

Large-scale assessments can be used to support learning by identifying areas in which students struggle and drawing attention and resources to these areas.[64] and to evaluate and adjust policies.[65] For example, PASEC results were used to shed light on issues of student retention in Senegal and demonstrate that holding back students did not result in improved learning outcomes the following year. In Madagascar, PASEC results showed that continuous teacher professional development was associated with lower learning outcomes. Further investigation revealed that this was because professional development was conducted during class time which reduced time teaching, and the policy was changed.[66]

While there is evidence that governments use the results of large-scale assessments to make policy changes, studies nevertheless question the impact of international large-scale assessments. For example, results from a 2017 survey suggest the results from international and regional assessments are more often being used to legitimize reforms that are already underway.[67] The linkages between participating in large-scale assessments and the impacts on policy and practice in DCPs need to be explored further. There are anecdotes showing how large-scale assessment can be used to improve learning, but the evidence is not packaged for policymakers to use. Guidelines and examples of how to use various assessments to improve learning in the classroom would fill a much-needed gap in global goods. Stakeholders consulted for this paper suggested that peer learning on how to use the results of learning assessment could be especially helpful for policy makers, as could an international expert group that makes its members available to provide technical assistance on using evidence for decision making.

Finally, innovation is also needed so that learning assessment data is used systematically to test interventions and reforms. To this end, a systematic test-intervene-retest approach using learning assessments could shed light on how countries can incorporate these assessments in their education programming, and find new ways of communicating and visualizing data to improve how learning assessment data are used. This test-intervene-retest approach could be piloted and documented in a small number of countries with the results made available as a global good.

KIX Opportunities and the GPE model

There are two ways in which the recommendations in this paper intersect with GPE's strategic plan and results framework. The first is that they addresses the greatest needs identified by DCPs and members of the GPE Partnership in support of improved and more equitable learning outcomes (strategic goal 1). As

[64] Kelleghan, Thomas, Vincent Greaney, and T. Scott Murray. *Using the Results of a National Assessment of Educational Achievement. Vol. 5*. Washington, DC: World Bank, 2009.
[65] UNESCO Institute for Statistics. "Learning Assessment Capacity Index." Montreal: UNESCO UIS. Accessed May 8, 2018.
[66] Bernard, Jean-Marc, and Katharina Michaelowa. "How Can Countries Use Cross-national Research Results to Address 'the Big Policy Issues?' (Case Studies from Francophone Africa).'" In Cross-national Studies of the Quality of Education: Planning Their Design and Managing Their Impact, 229-40. Paris: UNESCO IIEP, 2006.
[67] Fischman, Gustavo E., Amelia Marcetti Topper, Iveta Silova, Janna Goebel, and Jessica L. Holloway. "An Examination of the Influence of International Large Scale Assessments and Global Learning Metrics on National School Reform Policies." Journal of Education Policy, May 7, 2018.

a multi-stakeholder partnership, GPE is strategically positioned to take stock of the needs in this thematic area and create the space for those needs to be addressed. Not surprisingly, the themes emerging through the consultation process also overlap with several indicators in the results framework.

Lack of capacity at the country level is a major challenge for many countries and an impediment to monitoring learning outcomes, which is essential for GPE's funding model to be effective. GPE requires DCPs to have a system or mechanism to monitor learning outcomes or a time-bound plan to develop one. This requirement means that countries must have quality learning assessment systems, as defined under indicator 15 of the results framework (proportion of DCPs with a learning assessment system within the basic education cycle that meets quality standards). All of the proposed KIX opportunity areas discussed in this section can contribute to positive results on this indicator and thereby ensure that DCPs can monitor learning outcomes. Indicator 1 of the results framework on improved learning outcomes cannot be met without systems that can generate the required data on learning. In other words, learning cannot be improved unless progress and regressions are measured. The proposed KIX opportunity areas aim to build or strengthen systems to assess this learning.

To ensure that KIX investments in these opportunities respond to the current needs and landscape, three areas of investment are needed:

- Where there is a sufficient evidence base, capacity needs to be built through knowledge transfer, capacity development and learning exchange.
- Where there are some examples of solutions, but more synthesis is needed to develop a solid evidence base, it is necessary to build evidence and evaluation of what works.
- Where there is a need for new thinking and solutions, innovations in learning assessment systems are needed.

Examples of activities that could be supported through KIX

Table 2 shows exemplar activities that could be funded by KIX thematic funding for strengthening learning assessment systems in the potential investments just discussed. The dotted lines between the exemplar activities indicate that a recipient of KIX funding could potentially propose a project that covers multiple investments or opportunities.

Table 2: Areas of KIX Investment and Exemplar Activities

Areas of investment	Opportunities and Exemplar Activities			
	Global goods that support national learning assessment institutions	Global goods that build expertise in learning assessment throughout the education system	Global goods that support learning assessment systems for the most marginalized	Global goods that support the systemic use of learning assessment data
Building national capacity	Cross-country capacity building on designing aligned learning assessment frameworks Regional exchange on institution building	Building capacity among teachers for classroom-based assessments Building technical capacity of ministry officials	Participation of fragile and conflict-affected countries (FCACs) in cross-national assessments Guidelines for harmonizing donor priorities for assessment	Regional and international exchanges on good practices in using learning assessment data
Building evidence and evaluation of what works	In-depth analysis tools to diagnose learning assessment systems and recommend improvements Review of what works in public examinations reform Review of competency-based assessment initiatives	Review of good practices in classroom-based assessments	Review of the value for money of assessments Review of FCAC participation in national and cross-national assessments	Review of how learning assessment data (classroom-based, examinations, and large-scale) influence policy
Innovation	Groups of developing country partners co-developing assessments of skills beyond literacy and numeracy	Low-cost innovations in building sustainable expertise among ministries of education and teachers	Mechanisms for using existing data for global reporting Low-cost technology-based solutions for collecting data in FCACs and for students with disabilities and displaced learners	Systematic testing, intervening and retesting New ways of disseminating, communicating and visualizing learning assessment data

Bibliography

Anderson, Kate, Seamus Hegarty, Martin Henry, Helyn Kim, and Esther Care. *Breadth of Learning Opportunities: A Fresh Approach to Evaluating Education Systems*. Washington DC: Brookings Institution, 2018.

Anderson, Kate, and Joshua Muskin. *Learning Champions: How 15 Countries, Cities, and Provinces Came Together to Rethink Learning Assessment.* Washington DC: Brookings Institution, 2018.

Archer, Elizabeth. "The Assessment Purpose Triangle: Balancing the Purposes of Educational Assessment." *Frontiers in Education.* 2 (41), 2017.

Banerji, Rukmini, and Michael Walton. "What Helps Children to Learn? Evaluation of Pratham's Read India Program in Bihar & Uttarakhand." New Delhi: Pratham, 2011.

Bernard, Jean-Marc, and Katharina Michaelowa. "How Can Countries Use Cross-national Research Results to Address 'the Big Policy Issues?' (Case Studies from Francophone Africa).'" In *Cross-national Studies of the Quality of Education: Planning Their Design and Managing Their Impact*, 229-40. Paris: UNESCO IIEP, 2006.

Best, Maura, Pat Knight, Petra Leitz, Craig Lockwood, Dita Nugroho, and Mollie Tobin. "The Impact of National and International Assessment Programmes on Education Policy, Particularly Policies regarding Resource Allocation and Teaching and Learning Practices in Developing Countries." Melbourne: ACER, 2013.

Black, Paul, and Dylan Wiliam. "Developing a Theory of Formative Assessment." In J. Gardner, John, ed. *Assessment and Learning.* London: SAGE, 2006.

Burdett, Newman. "Review of High Stakes Examination Instruments in Primary and Secondary School in Developing Countries." RISE Working Paper 17/018, 2017.

Care, Esther, Helyn Kim, Kate Anderson, and Emily Gustafsson-Wright. *Skills for a Changing World: National Perspectives and the Global Movement.* Washington DC: Brookings Institution, 2017.

Care, Esther, Alvin Vista, and Helyn Kim. "Optimizing Assessment for All." Washington, DC: Brookings Institution., 2018.

Chinapah, Vinayagum. *Monitoring Learning Achievement: Toward Capacity Building.* Paris: UNESCO, 1997.

Clarke, Marguerite. "What Matters Most for Student Assessment Systems: A Framework Paper." Washington DC: World Bank, 2012.

Fischman, Gustavo E., Amelia Marcetti Topper, Iveta Silova, Janna Goebel, and Jessica L. Holloway. "An Examination of the Influence of International Large Scale Assessments and Global Learning Metrics on National School Reform Policies." *Journal of Education Policy*, May 7, 2018.

Friedman, Tim, Ursula Schwantner, Jeaniene Spink, Naoko Tabata, and Charlotte Waters. *Improving Quality Education and Children's Learning Outcomes and Effective Practices in the Eastern and Southern Africa Region.* Nairobi: UNICEF ESARO, 2016.

Global Partnership for Education. 2016. "GPE 2020 Strategic Plan." Global Partnership for Education. Washington DC: GPE, 2016.

———. "GPE Results Framework for 2016-2020." Global Partnership for Education. Washington DC: GPE, 2017.

Greaney, Vincent, and Thomas Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Washington, DC: World Bank, 2008.

High-Level Panel of Eminent Persons on the Post-2015 Development Agenda. *A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development*. New York: United Nations, 2013.

IEA. "IEA Studies." Accessed May 08, 2018.

The International Commission on Financing Global Education Opportunity. *The learning generation: Investing in education for a changing world.* Washington DC: Education Commission, 2016.

Kanjee, Anil. Excerpt from consultation feedback. 10 August 2018.

Kelleghan, Thomas, Vincent Greaney, and T. Scott Murray. *Using the Results of a National Assessment of Educational Achievement*. Vol. 5. Washington, DC: World Bank, 2009.

Learning Metrics Task Force. *Toward Universal Learning: Recommendations from the Learning Metrics Task Force*. Montreal and Washington DC: UIS and Brookings, 2013.

———. *Toward Universal Learning: Implementing Assessment to Improve Learning.* Montreal and Washington DC: UIS and Brookings, 2014.

Lockheed, Marlaine E. *The Craft of Education Assessment: Does Participating in International and Regional Assessments Build Assessment Capacity in Developing Countries? An Independent Evaluation of IEA's Program on the Assessment of Student Achievement (PASA),* 2010.

Mugo, John. "Welcoming Two New Initiatives to Accelerate Quality Teaching and Learning in Africa." Washington DC: Brookings Institution, 2017.

Mutambo, Aggrey. "UK Company Fined Sh330 Million for Bribing Kenya Officials." *Daily Nation*. January 11, 2016.

OECD. *Building Capacity for Assessment in PISA for Development Countries.* Paris: OECD, 2017.

Petters, Janet Sunday, and Maureen O. Okon. "Students' Perception of Causes and Effects of Examination Malpractice in the Nigerian Educational System: The Way Forward for Quality Education." *Procedia - Social and Behavioral Sciences* 114: 125-29, 2014.

Pritchett, Lant. “Creating Education Systems Coherent for Learning Outcomes: Making the Transition from Schooling to Learning.” London: RISE Programme, 2014.

Ramirez, Maria Jose. *Quick Guide No. 3 Implementing a National Learning Assessment.* Montreal: UIS, 2017. http://uis.unesco.org/sites/default/files/documents/quick-guide-3-implementing-national-learning-assessment.pdf.

Save the Children. *Beyond School Walls: A Boost for Readers*. Fairfield CT: Save the Children, 2013.

SDSN Thematic Group 4. *The Future of Our Children: Lifelong, Multi-Generational Learning For Sustainable Development.* New York: SDSN, 2014.

Somerset, H.C.A. *Examination Reform in Kenya*. Washington DC: World Bank, 1987.

Tobin, Mollie, Petra Lietz, Dita Nugroho, Ramya Vivekanandan, and Tserennadmid Nyamkhuu. *Using Large-scale Assessments of Students Learning to Inform Education Policy: Insights from the Asia-Pacific Region*. Melbourne and Bangkok: ACER and UNESCO Bangkok, 2015.

Transparency International. *Global Corruption Report: Education*. London: Earthscan, Publishing for a Sustainable Future, Routledge, Taylor & Francis Group, 2013.

UNESCO. “Global Education Monitoring Report 2010: Reaching the Marginalized.” Paris: UNESCO, 2010.

———. “Policy paper 34: Fulfilling our collective responsibility: Financing global public goods in education.” Paris: UNESCO, 2018.

UNESCO Institute for Statistics. “Learning Assessment Capacity Index.” Montreal: UNESCO UIS. Accessed May 8, 2018.

———. “6 Out of 10 Children and Adolescents Are Not Learning a Minimum in Reading and Math.” Montreal: UNESCO UIS, 2017a.

———. “Investment Case for Expanding Coverage and Comparability for Global Indicator 4.1.1.” Montreal: UNESCO UIS, 2017b.

UIS (UNESCO Institute for Statistics) and ACER (Australian Council for Educational Research). *Principles of Good Practice in Learning Assessment*. Camberwell, Victoria: ACER, 2017.

UNESCO International Institute for Capacity Building in Africa. *Teaching policies and learning outcomes in Sub-Saharan Africa: Issues and Options.* Addis Ababa: UNESCO IICBA, 2016.

USAID. “Early Grade Reading Barometer.” Accessed May 8, 2018. http://www.earlygradereadingbarometer.org/.

Wagner, Daniel, Sharon Wolf and Robert F. Boruch, eds. *Learning at the Bottom of the Pyramid: Science, Measurement and Policy in Low-Income Countries.* Paris: UNESCO IIEP, 2018.

Wagner, Daniel, Martin Lockheed, Ina Mullis, Michael O. Martin, Anil Kanjee, Amber Gove, and Amy Dowd. "The Debate on Learning Assessments in Developing Countries." *Journal of Comparative and International Education*, *42*(3):509-545, 2012.

Winthrop, Rebecca, Eileen McGivney, and Adam Barton. *Can We Leapfrog? The Potential of Education Innovations to Rapidly Accelerate Progress.* Washington DC: Brookings Institution, 2017.

World Bank. "Student Assessment for Policymakers and Practitioners | World Bank Group." Accessed May 08, 2018. https://olc.worldbank.org/content/student-assessment-policymakers-and-practitioners.

Annex A: GPE Fact Sheet on Learning Assessments

Overview of Indicators 1, 15 and 20
Three results framework indicators are related to learning assessments:

- Indicator 1: proportion of developing country partners (DCPs) showing improvement on learning outcomes
- Indicator 15: proportion of DCPs with a learning assessment system within the basic education cycle that meets quality standards
- Indicator 20: proportion of grants supporting education management and information systems (EMIS)/learning assessment systems.[1]

Overall, DCPs face challenges related to the availability of learning assessments to track progress in learning outcomes (indicator 1). Only 32 percent (19 out of 60 DCPs assessed) had learning assessment systems that met quality standards between 2011 and 2015 (indicator 15). To tackle this measurement gap, GPE is integrally engaged in supporting learning assessments in DCPs through Education Sector Program Implementation Grants (ESPIGs) (indicator 20).

Key results
The first milestone for improving learning outcomes (indicator 1) is set for 2018. [2] The target is 68 percent of all DCPs showing improvement on learning outcomes, including 65 percent in fragile and conflict-affected countries (FCACs). The 2016 Results Report noted that learning outcomes had improved in 13 out of 20 developing country partners with available data at two points in time (two out of four FCACs) at baseline, between 2000 and 2015.

Baseline data for indicator 15 suggest that fewer than one in three developing country partners (32 percent) had learning assessment systems (21 percent in FCACs) that met quality standards.[3] The next milestone for indicator 15 has been set for 2018, with a target of 38 percent of all DCPs having a learning assessment system that meets quality standards, including 29 percent in FCACs. No new data are available for 2017 for this indicator.

No milestone is set for the proportion of ESPIGs supporting EMIS/learning assessment systems (indicator 20) for 2017.[4] However, data show that 92 percent (44 out of 48) ESPIGs were active in FY2017, supported EMIS/learning assessment systems, far exceeding the indicator's first milestone target set for 50 percent in 2018.[5] Among FCACs, this figure was 96 percent (26 out of 27), again much higher than the 44 percent milestone target set for 2018.

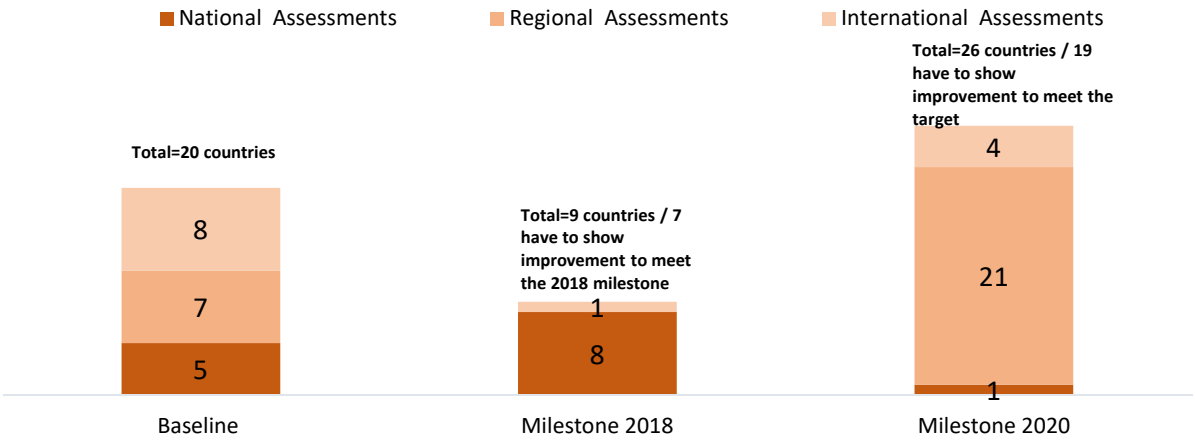
[1] There is no breakdown by gender for these three indicators in the results framework.
[2] For indicator 1's methodology sheet, see https://www.globalpartnership.org/content/methodology-sheet-gpe-result-indicator-1.
[3] For indicator 15's methodology sheet, which draws from the SABER Student Assessment module but is further contextualized for GPE's DCPs, see https://www.globalpartnership.org/content/methodology-sheet-gpe-result-indicator-15.
[4] For indicator 20's methodology sheet, seehttps://www.globalpartnership.org/content/methodology-sheet-gpe-result-indicator-20.
[5] This includes seven pooled-funded Education Sector Program Implementation Grants.

Progress in learning assessments

Although there was no milestone target for indicator 1 in 2017, GPE tracks the availability of learning assessment data. Several international, regional and national assessments will have been implemented between 2011 and 2019 in DCPs, which will enable GPE to calculate and report on improvements in learning. GPE estimates nine countries will have participated twice in the same assessments between 2011 and 2017 and 26 between 2011 and 2019 (Figure 1).

Figure 1: Data to evaluate learning progress will increase slightly by 2020

Number of countries with available learning assessment data to measure progress in learning outcomes in 2018 and 2020



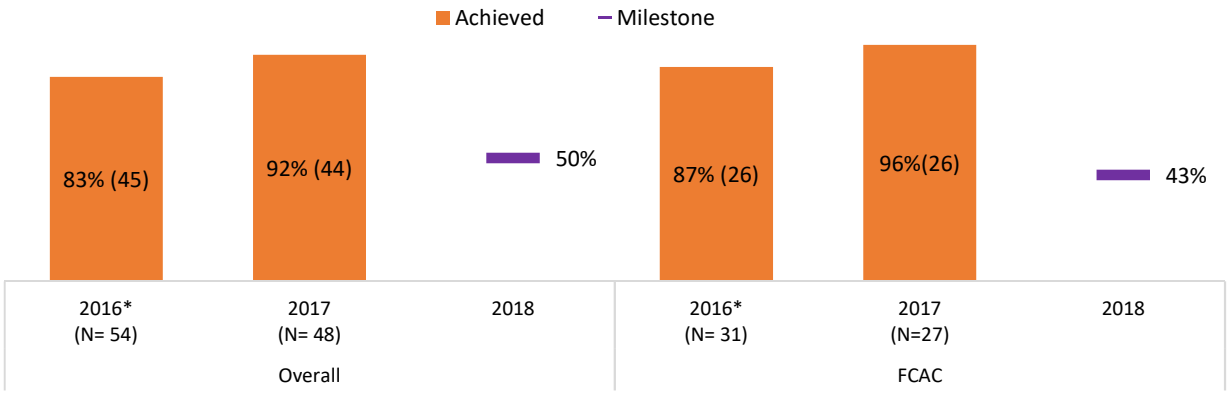
GPE is strengthening its support for learning assessment as demonstrated by the increase in the proportion of ESPIGs supporting learning assessment systems (LAS) and/or EMIS. As mentioned above, current data on Indicator 20 show that 92 percent (44 out of 48) of ESPIGs[6] active in FY2017 supported EMIS and/or LAS, up from 83 percent (45 out of 54) in FY2016 (Figure 2).[7]

[6] This includes seven pooled-funded ESPIGs.

[7] The methodology for indicator 20 changed in FY2017. The FY2016 data was obtained by re-coding FY2016 data using the same methodology used for FY2017.

Figure 2: Almost all GPE grants support EMIS and/or LAS

Proportion of active grants supporting EMIS and/or LAS, FY2016 and FY2017



Note: FY16 figures are based on re-coding.

Annex B. National and Cross-National Assessments in Developing Country Partners, 2010–2017

Country	Types of Learning Assessments Conducted by Year				
	National	Regional	International	Citizen-led	Early Grade
Afghanistan	2013				EGRA 2015
Albania	2015		PISA 2012, 2015		
Bangladesh	2013			2015	
Benin	2011	PASEC 2014			EGRA 2015
Bhutan	2013				
Burkina Faso	2014	PASEC 2014			
Burundi	2011	PASEC 2014			EGRA 2011
Cambodia	2011	PASEC 2014 *SEA-PLM 2016			EGRA 2016
Cameroon	2011	PASEC 2014			EGRA 2015
Central African Republic					
Chad		PASEC 2014			EGRA 2012
Comoros					
Congo, Democratic Republic of					EGRA 2015, EGMA 2015
Congo, Republic of		PASEC 2014			
Côte d'Ivoire	2011	PASEC 2014			
Djibouti					
Dominica					
Eritrea	2015				
Ethiopia	2013				EGRA 2014
The Gambia	2014				EGRA 2013
Georgia	2015		PIRLS 2010, 2015 TIMSS 2010, 2015		
Ghana	2014		TIMSS 2010	2016	EGRA 2015, EGMA 2013
Grenada					
Guinea	2014				
Guinea-Bissau	2014				
Guyana	2014				
Haiti			*LaNA 2016		EGRA 2015

Country	Types of Learning Assessments Conducted by Year				
	National	Regional	International	Citizen-led	Early Grade
Honduras	2014	TERCE 2013	PIRLS 2010 TIMSS 2010		EGRA 2016
Kenya		SEACMEQ 2013		2015	EGRA 2015 EGMA 2012
Kyrgyz Republic			PISA 2015		EGRA 2011
Lao People's Democratic Republic	2012	PASEC 2014 *SEA-PLM 2016			EGRA 2016
Lesotho	2014	SEACMEQ 2013			EGRA 2015
Liberia					EGRA 2015 EGMA 2015
Madagascar	2012				
Malawi	2012	SEACMEQ 2013			EGRA 2016 EGMA 2011
Mali		PASEC 2014		2016	EGRA 2015, EGMA 2011
Mauritania					
Moldova			PISA 2015		
Mongolia	2013				
Mozambique	2013	SEACMEQ 2013		2016	EGRA 2015
Nepal	2013				EGRA 2014
Nicaragua	2010	TERCE 2013			EGRA 2015 EGMA 2011
Niger		PASEC 2014			EGRA 2014
Nigeria	2011				EGRA 2015 EGMA 2013
Pakistan	2014			2016	EGRA 2014
Papua New Guinea		PILNA 2012			EGRA 2013
Rwanda	2011				EGRA 2015 EGMA 2011
Saint Lucia					
Saint Vincent and the Grenadines					
São Tomé and Príncipe	2015				
Senegal	2014	PASEC 2014		2013	EGRA 2014
Sierra Leone			MICS 2018		EGRA 2012 EGMA 2014
Somalia					EGRA 2013

Country	Types of Learning Assessments Conducted by Year				
	National	Regional	International	Citizen-led	Early Grade
South Sudan					EGRA 2013
Sudan	2015				EGRA 2015
Tajikistan					EGRA 2016
Tanzania	2014	SEACMEQ 2013		2012	EGRA 2016 EGMA 2016
Timor-Leste					EGRA 2011
Togo	2013	PASEC 2014	MICS 2018		EGRA 2015
Uganda	2013	SEACMEQ 2013		2015	EGRA 2015
Uzbekistan					
Vietnam	2013	PASEC *SEA-PLM 2016	PISA 2012, 2015		EGRA 2010
Yemen			TIMSS 2010		
Zambia	2013	SEACMEQ 2013			EGRA 2016 EGMA 2011
Zimbabwe	2014	SEACMEQ 2013			EGRA 2013
Total (65)	**39**	**25**	**11**	**11**	**41**

* = field trial only; EGRA = early grade reading assessment; EGMA = early grade mathematics assessment; LaNA = literacy and numeracy assessment; MICS = multiple indicator cluster survey foundational learning skills module; PASEC = Programme d'analyse des systèmes éducatifs de la CONFEMEN; PILNA = Pacific Islands Literacy and Numeracy Assessment; PIRLS = Progress in International Reading Literacy Study; PISA = Programme for International Student Assessment; SEA-PLM = Southeast Asia primary learning metric; SEACMEQ = Southern and Eastern Africa Consortium for Monitoring Educational Quality; TERCE = Tercer Estudio Regional Comparativo y Explicativo; TIMSS = Trends in International Mathematics and Science Study.
Note: DCPs that have not conducted the listed learning assessments during this timeframe are indicated in blue below.
Source: UIS Catalogue of Learning Assessment and Learning Assessment Capacity Index; PAL Network, early grade reading assessment tracker and early grade mathematics tracker (RTI).