

METHODOLOGY SHEET FOR GLOBAL PARTNERSHIP FOR EDUCATION (GPE) INDICATORS

title	Indicator (15) Proportion of DCPs with a learning assessment system within the basic education cycle that meets quality standards
Result measured (from GPE Results Framework):	Strategic Goal (3) Effective and efficient education systems delivering equitable, quality educational services for all
JUSTIFICATION FOR INDICATOR	
Background/context for indicator:	<p>More children than ever are in school, but some reports indicate that the quality of education in many developing countries has dropped in recent years¹. The benefits of education – for national development, individual prosperity, health and social stability – are well known, but for these benefits to accrue, children in school have to be learning. Despite commitments and progress made in improving access to education at the global level, including Millennium Development Goal (MDG) 2 on universal primary education and the Education for All (EFA) goals, levels of learning are still too low. Many children and youth complete primary and secondary education without acquiring the basic knowledge, skills and competencies they need to lead productive, healthy lives².</p> <p>To reform this crisis in education, learning assessments are increasingly gaining recognition as an integral part of successful education systems across the world. The importance of high quality assessments has been widely acknowledged. Assessment is believed to lie at the heart of schooling, alongside curriculum, learning, and teaching³. The Global Partnership for Education (GPE) is dedicated to supporting their country partners’ commitment to monitor learning outcomes. GPE Secretariat worked with The Education Policy and Data Center (EPDC) at FHI 360 to develop a methodology that can be used to assess the quality of learning assessment systems in GPE Developing Country Partners (DCPs)⁴.</p>
Rationale for indicator selection:	GPE has been working to help countries use existing or new learning assessments to monitor progress and improve the reliability of data used for policy making. It is included in the data requirement of GPE funding model: “Countries applying for a Program Implementation Grant must have a system or mechanisms to monitor learning outcomes, such as a national learning assessment system, OR a time-bound plan to develop mechanisms to monitor learning outcomes”.

¹ D Wagner, *Smaller, Quicker, Cheaper: Improving Learning Assessments for Developing Countries* (2011), retrieved from: <http://unesdoc.unesco.org/images/0021/002136/213663e.pdf>

² Learning Metrics Task Force, *Toward Universal Learning- Implementing Assessment to Improve Learning* (2014), retrieved from: <http://www.uis.unesco.org/Education/Documents/towards-universal-learning-a-global-framework-for-measuring-learning-metrics-task-force-education-2013-en.pdf>

³ P Hill and M Barber, *Preparing for a Renaissance in Assessment* (2014), retrieved from: https://uk.pearson.com/content/dam/region-core/uk/pearson-uk/documents/educator/primary/preparing_for_a_renaissance_in_assessment_and_summary_text_december_2014.pdf

⁴ Education Policy Data Center, FHI 360, *National Learning Assessment Mapping Project* (2015) retrieved from: http://www.epdc.org/sites/default/files/documents/EPDC_NLAMP_report-v3.pdf

Assessment regimes are a central part of the reform effort to fix the learning crisis around the world. They can provide evidence on the scale of the learning crisis as a lever for reform. They can also track progress on improvements and provide the evidence base for what works and can empower parents to demand better outcomes—or move their kids to where they can find them⁵. This indicator can serve the GPE to track assessment systems and better support countries in the strengthening of such systems. This will in turn empower countries to build capacity and enhance technical abilities of personnel. Effective learning is the ultimate goal of education systems, and this indicator will help inform areas where systemic change is essential.

DEFINITION

Indicator definition:

The definition and criteria for assessing the quality of learning assessment systems are guided by the framework for learning assessments developed by the World Bank's Systems Approach for better Education Results (SABER). SABER defines an **assessment** 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*, while an **assessment system** is described as *a group of policies, structures, practices, and tools for generating and using information on student learning and achievement*⁶.

Three determinants taken from SABER are used to identify the quality of learning assessment systems in this methodology. However, we developed our own criteria to assess each of these determinants.

1. **Enabling Context**- Context in which an assessment activity takes place. This determinant specifically looks at the frequency of assessments being offered, what it measures and the levels at which it was conducted. Generally, a permanent agency, institution or an office is responsible for conducting this assessment within the country.
2. **Assessment quality**- This is related to the instruments, processes and procedures used. Indicators selected to measure the **quality of assessment** focus on the technical strength of the assessment activities and the transparency of results. The proposed methodology assumes that a technical/methodological document is available to the general public that describes the assessment in detail and that the assessment results are widely shared with the general public within 12 months of data collection. It assumes that international assessments (such as PISA, TIMSS, and PIRLS)⁷ and regional assessments (such as SACMEQ, PASEC, and LLECE)⁸ are transparent with results and applies this criterion of reporting within 12 months only to examinations and national large-scale assessments (NLSA).

System Alignment- This determinant helps assess whether the learning assessment is based on official learning standards and/or curriculum. It assumes that examinations are by default based on the curriculum and applies the curriculum alignment criterion only to NLSAs, in large part relying on available methodology documents to indicate that. The curriculum criterion is not considered applicable for international or regional large-scale assessments.

⁵ R Banerji and L Pritchett, *Schooling is Not Education! Using assessment to change the politics of non-learning* (2013), retrieved from: http://www.cgdev.org/doc/full_text/CGDReports/3120290/schooling-is-not-education.html

⁶ M Clarke, *What matters most for student assessment systems: a framework paper (Report No. 68235)* (2012), 6, retrieved from: <https://openknowledge.worldbank.org/bitstream/handle/10986/17471/682350WP00PUBL0WP10READ0web04019012.pdf;sequence=1>

⁷ Trends in International Mathematics and Science Study, Progress in International Reading Literacy Study, Programme for International Student Assessment

⁸ The Southern and Eastern Africa Consortium for Monitoring Educational Quality, Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (Conférence des Ministres de l'Éducation des pays africains et malgache d'expression française), Latin American Laboratory for Assessment of the Quality of Education

Each of the three determinants (enabling context, assessment quality, and system alignment) is applied to all large-scale learning assessments (including national, regional, and international assessments) and examinations within each country⁹. To reflect the different nature of examinations and large scale assessments, criteria across the three determinants used to assess them are distinct. The criteria developed for examinations focus on their frequency, i.e. whether they are conducted on at least an annual basis, and their inclusiveness, i.e. whether they are offered to all eligible individuals without discrimination. For large-scale assessments, the criteria focus on whether an assessment has been conducted at least twice, and whether the most recent data collection effort has taken place between 2011 and 2015. For both examinations and large-scale assessments, an important criterion to consider is whether there is a permanent agency or institution conducting the assessment. In sum, these criteria are selected to indicate the degree to which an assessment is established enough to produce results that can be used for monitoring learning over time, as opposed to a one-off assessment that can generate “shock value” but cannot contribute to better understanding of changes in learning achievement⁶.

The table below demonstrates all the criteria used for this exercise. A learning assessment system will have met the quality standards if it is classified as established. See sections below for an explanation of which criteria need to be met for the learning assessment system to be classified as such.

Table 1: Criteria guiding the data collection process

		Examinations	Large scale learning assessment
ENABLING CONTEXT			
The assessment is offered:	i) annually or more frequently	yes/no	N/A
	ii) to all eligible individuals	yes/no	N/A
The assessment has been carried out:	i) at least twice at the basic education level	N/A	yes/no
	ii) most recently between 2011 and 2015	N/A	yes/no
Assessment measures:	i) language	yes/no	yes/no
	ii) math	yes/no	yes/no
	iii) science	yes/no	yes/no
	iv) other subjects	yes/no	yes/no
Assessment is available at levels:	i) primary	yes/no	yes/no
	ii) lower secondary	yes/no	yes/no
A permanent agency, institution, or office conducts the assessment		yes/no	yes/no
SYSTEM ALIGNEMENT			
The assessment is based on official learning standards/curriculum		N/A	yes/no ¹⁰

⁹ Early Grade Reading Assessment (EGRA) results have been used when the sampling and comparability of the results have allowed for national monitoring of learning outcomes. Citizenship-led assessments have been excluded.

¹⁰ This criteria is not applicable to international or regional large scale assessments.

ASSESSMENT QUALITY			
	There is a publicly available methodology/technical document on the assessment	yes/no	yes/no
	The assessment results are made available to the general public within 12 months	yes/no	yes/no ¹¹
Unit of measurement:	<p>"n out of N", expressed as a percentage, where:</p> <ul style="list-style-type: none"> - n refers to the number of DCPs with a learning assessment system that meets quality standards; that is, a LAS that meets the respective criteria for being "established" as determined through the GPE methodology - N refers to the total number of DCPs in the sample 		
Disaggregation:	By Fragile and Conflict-affected Country / Non-fragile and Conflict-affected Country status		
Year for data reported (select only one and mark an "x")	___ fiscal year _X_ calendar year		
Frequency of data collection:	The corporate indicator is updated every two years, as learning assessment systems are not expected to change on a yearly basis. That is, data will be collected in 2018 and in 2020 for the purpose of the GPE results framework under the GPE 2020 strategic plan.		

DATA TREATMENT

Source of information for collecting data:	Source document, template, etc.:	<ol style="list-style-type: none"> 1. UIS Learning Assessment Catalogue 2. UIS Central Data Catalogue 3. Education assessment reports put out by relevant ministries for countries under consideration 4. SABER Country Reports 5. Any other country-level data sources specific to the DCP
	Source agency:	<ol style="list-style-type: none"> 1. UIS 2. UNESCO 3. World Bank 4. DCP's Ministries of Education
Formula:	<p>Country Sample: The country sample includes 60 GPE development country partners, excluding the 5 countries that joined in April 2016, after completing the data collection for this indicator. The total number of countries will stay the same for comparability between years.</p> <p>The methodology used to calculate the indicator consists of a 5-step process, where criteria are first scored, examinations are classified, large-scale assessments are classified, an overall classification for the learning assessment system is derived, and the number of countries that have met the quality standard of "established" are counted. Below is an explanation of these steps.</p> <p>Step 1: Assign a score to each criterion</p> <ul style="list-style-type: none"> • Examinations and Large Scale Learning Assessments (national, regional, and international) are assessed for each of the 60 countries using criteria corresponding to the three SABER determinants used in this methodology (i.e. enabling context, system alignment, and assessment quality). • Criteria for each of these determinants are scored based on publicly 	

¹¹ This criteria is not applicable to international or regional large scale assessments.

available information.

- Responses are computed as follows, for each criterion:

1: Yes

0: No

N/A: Not applicable

Some criteria are not applicable for examinations and others for large scale assessments. Those criteria are not scored (see Table 1 in this methodology sheet for the criteria included in each of the determinants).

Step 2: Calculate the weighted value for each criterion

- Different weights have been assigned across the criteria, across the three determinants. The following two criteria (1) the number of subjects in which examinations and NLSA take place and (2) education levels (just for examinations) are important, but not as crucial as the assessment practices themselves taking place on a regular basis (annually or more frequently). Hence for these two criteria a score of at least 1 across its sub criteria gives the criterion a score of 1. If either of these two criteria have a total of 2 or above across sub criteria, the score remains 1.
- All the other criteria are weighted equally and assigned a value of 1 for every positive response.

Step 3: Assigning a category (“no information,” “nascent,” “under development,” or “established”) for examinations and large scale assessments within a country

- Once all the criteria are assigned weighted values, the scores are summed up for examinations and each individual large scale assessment separately, based on the weighting mentioned in Step 2.
- **For examinations**, the sum of weighted values for specific criteria are used to determine an overall classification. Categories are assigned as follows:

- **No information**: examinations are classified as “no information” when no data was found or was publicly available (sum of weighted values = 0)

- **Nascent**: examinations are classified as “nascent” when it meets some but not all of the following criteria- a) examination is offered annually, b) to all eligible students, c) for at least one subject, and d) at least at one level of basic education (sum of weighted values for the specific criteria = 1-3).

- **Under development**: examinations are classified as “under development” when it meets all of the following criteria- a) examination is offered annually, b) to all eligible students, c) for at least one subject, and d) at least at one level of basic education, but does not meet the additional set of criteria to be “established” (sum of weighted values for the specific criteria = 4-5).

- **Established**: examinations are classified as “established” when it meets all of the criteria required to be classified as “under development” in addition to the following criteria- a) there is permanent agency/institution/office that conducts the assessment, and b) there is a publicly available methodology/technical document on the assessment or assessment results are made available to the general public within 12 months (sum of weighted values for specific criteria = 6-8).

Table 2 summarizes these details.

Table 2: Criteria required for examinations to be categorized as “under development” or “established”

		Examinations
Criteria required to fall under the “under development” category:		
The assessment is offered:	i) annually or more frequently	X
	ii) to all eligible individuals	X
The assessment is offered for at least one subject (i.e., language, mathematics, science, and/or other subjects)		X
The assessment is offered for at least one level of basic education (i.e., primary and/or lower secondary education)		X
Supplementary criteria required to fall under the “established” category (in addition to the criteria required for the “under development” category as listed above):		
A permanent agency, institution, or office conducts the assessment		X
There is a publicly available methodology/technical document on the assessment OR the assessment results are made available to the general public within 12 months		X

- **For large scale assessments**, the sum of weighted values for specific criteria are used to determine a classification for each individual large scale assessment that took place (national, regional, international). Categories for each available assessment are assigned as follows:

- **No information**: a large-scale assessment is classified as “no information” when no data was found or was publicly available (sum of weighted values = 0)

- **Nascent**: a large scale assessment is classified as “nascent” when it meets some but not all of the following criteria- a) the assessment has been carried out at least twice, b) most recently between 2011-2015, c) for at least one subject, d) at least at one level of basic education, and if it is a national large scale assessment e) is based on official learning standards/curricula (sum of weighted values for national large scale assessments=1-4, sum of weighted values for international/regional assessments = 1-3).

- **Under development**: a large-scale assessment is classified as “under development” when it meets all of the following criteria- a) the assessment has been carried out at least twice, b) most recently between 2011-2015, c) for at least one subject, d) at least at one level of basic education, and if it is a national large scale assessment e) is based on official learning standards/curricula. However, the assessment does not meet the additional set of criteria to be “established” (sum of weighted values for national large scale assessments = 5-6, sum of weighted values for international/regional assessments = 4-5).

- **Established**: a large scale assessment is classified as “established” when it meets all of the criteria required to be classified as “under development” in addition to the following criteria- a) there is permanent agency/institution/office that conducts the assessment, and b) there is a publicly available methodology/technical document on the assessment *or* if it is a national large scale assessment, assessment results are made available to the general public within 12 months (sum of weighted values for national large scale assessments = 7, sum of weighted values for international/regional assessments = 6).

To determine the overall classification for large scale assessments, the highest

category achieved by any of the large-scale assessments (national, regional, international) is retained.

Table 3 below summarizes these details.

Table 3: Criteria required for large scale assessments to be placed in the “under development” and “established” categories

		NLSAs	International/ regional LSAs
Criteria required to fall under the “under development” category:			
The assessment has been carried out:	i) at least twice at the basic education level	X	X
	ii) most recently between 2011 and 2015	X	X
The assessment is offered for at least one subject (i.e., language, mathematics, science, and/or other subjects)		X	X
The assessment is offered for at least one level of basic education (i.e., primary and/or lower secondary education)		X	X
The assessment is based on official learning standards/curriculum		X	N/A
Supplementary criteria required to fall under the “established” category (in addition to the criteria required for the “under development” category as listed above):			
A permanent agency, institution, or office conducts the assessment		X	X
There is a publicly available methodology/technical document on the assessment OR if it is a national large scale assessment, the assessment results are made available to the general public within 12 months.		X	X

Step 4: Assigning one single category for each country

Once the countries are assigned two categories, one for their examinations and another for their large-scale assessments, the learning assessment system of each country receives an overall classification according to the possible combinations demonstrated in Table 4 below:

Table 4: Categories of learning assessment systems, achieved by combining the classifications of examinations and LSAs.

Categories assigned to examination and LSA activities	Learning assessment system
No information + No information	No information
Nascent + Nascent	Nascent
No information + Nascent	
No information + Under development	
Nascent + Under development	
Under development + Under development	Under development
Nascent + Established	
Under Development + Established	
No information on Examination + Established on regional/international LSA	
No information on LSA + Established on Examination	
Established + Established	Established

No info on Examinations + Established on **national**
LSA

Step 5: Counting the number of countries within each categories and calculating the value for the corporate indicator

Once the categorization of learning assessment systems as a whole is finalized by country, the number of countries under each group is added up for reporting purposes. Countries with a learning assessment system that is classified as “established” meet the quality standard for the indicator. The number of countries that meet the quality standard is then divided by the total number of DCPs in the sample and multiplied by 100 to obtain the corporate level indicator.

Step 1 - 4 (corresponding to steps mentioned in ‘Calculation method’): Assigning a category for examinations and for LSA individually for each of the countries

The following equations are applied to every country in the list individually for Examinations and Learning Assessments, separately:

a. For Examinations

$$NoInfoExam_c = \begin{cases} 1, & \text{if no data found} \\ 0, & \text{otherwise} \end{cases}$$

$$NascentExam_c = \begin{cases} 1, & \text{if } UnderDevExam_c = 0 \text{ and } EstablishedExam_c = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$UnderDevExam_c = \begin{cases} 1, & \text{if } (Context1 = 1) \text{ and} \\ & (Context2 = 1) \text{ and} \\ & \left(\sum_{i=1}^n ContextSub > 0 \right) \text{ and} \\ & \left(\sum_{j=1}^n ContextLevel > 0 \right) \text{ and} \\ 0, & \text{otherwise} \end{cases}$$

Aggregation formula:

$$EstablishedExam_c = \begin{cases} 1, & \text{if } (UnderDevExam_c) \text{ and} \\ & (Agency = 1) \text{ and} \\ & (Quality1 = 1) \text{ or} \\ & (Quality2 = 1) \\ 0, & \text{otherwise} \end{cases}$$

where,

i= subject 1... country n (i.e. Math, Language, science and other subjects)

j= education level1... education level n (i.e. Primary & Secondary)

c= country 1... country n (i.e. 60 GPE member countries in the sample)

Context1= Assessment is offered annually or more frequently

Context2= Assessment is offered to all eligible individuals

ContextSub= The assessment has been offered for at least one subject (i.e. math, science, language, or other)

ContextLevel= The assessment is available at i: primary ii: secondary levels

Agency= A permanent agency, institution, or office conducts the assessments

Quality1= There is a publicly available methodology/technical document on the assessment

Quality2= The assessment results are made available to the general public within 12 months

b. For National Large Scale Learning Assessment

$$NoInfoNLSA_c = \begin{cases} 1, & \text{if no data found} \\ 0, & \text{otherwise} \end{cases}$$

$$NascentNLSA_c = \begin{cases} 1, & \text{if } UnderDevNLSA_c = 0 \text{ and } EstablishedNLSA_c = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$UnderDevNLSA_c = \begin{cases} 1, & \text{if } (Context3 = 1) \text{ and} \\ & (Context4 = 1) \text{ and} \\ & (ContextStandard = 1) \text{ and} \\ & \left(\sum_{i=1}^n ContextSub > 0 \right) \text{ and} \\ & \left(\sum_{j=1}^n ContextLevel > 0 \right) \text{ and} \\ 0, & \text{otherwise} \end{cases}$$

$$EstablishedNLSA_c = \begin{cases} 1, & \text{if } (UnderDevNLSA_{c_{1...60}}) \text{ and} \\ & (Agency = 1) \text{ and} \\ & (Quality1 = 1) \text{ or} \\ & (Quality2 = 1) \\ 0, & \text{otherwise} \end{cases}$$

where,

i= subject 1... country n (i.e. Math, Language, science and other subjects)

j= education level1... education level n (i.e. Primary & Secondary)

c= country 1... country n (i.e. A 60 GPE member countries in the sample)

Context3= Assessment has been carried out at least twice at the basic education level

Context4= Assessment has been carried out most recently between 2011 and 2015

ContextStandard= The assessment is based on official learning standards/curriculum

ContextSub= The assessment has been offered for at least one subject (i.e. math, science, language, or other subject)

ContextLevel= The assessment is available at i: primary ii: secondary levels

Agency= A permanent agency, institution, or office conducts the assessments

Quality1= There is a publicly available methodology/technical document on the assessment

Quality2=The assessment results are made available to the general public within 12 months

c. For International or Regional Large Scale Learning Assessment

$$NoInfoILSA_c = \begin{cases} 1, & \text{if no data found} \\ 0, & \text{otherwise} \end{cases}$$

$$NascentILSA_c = \begin{cases} 1, & \text{if } UnderDevILSA_c = 0 \text{ and } EstablishedILSA_c = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$UnderDevILSA_c = \begin{cases} 1, \text{ if } (Context3 = 1) \text{ and} \\ \text{ } (Context4 = 1) \text{ and} \\ \left(\sum_{i=1}^n ContextSub > 0 \right) \text{ and} \\ \left(\sum_{j=1}^n ContextLevel > 0 \right) \text{ and} \\ 0, \text{ otherwise} \end{cases}$$

$$EstablishedILSA_c = \begin{cases} 1, \text{ if } (UnderDevILSA_{c_{1...60}}) \text{ and} \\ \text{ } (Agency = 1) \text{ and} \\ \text{ } (Quality1 = 1) \\ 0, \text{ otherwise} \end{cases}$$

where,

i= subject 1... country n (i.e. Math, Language, science and other subjects)

j= education level1... education level n (i.e. Primary & Secondary)

c= country 1... country n (i.e All 60 GPE member countries)

Context3= Assessment has been carried out at least twice at the basic education level

Context4= Assessment has been carried out most recently between 2011 and 2015

ContextSub= The assessment has been offered for at least one subject (ie: math, science, language, or other subject)

ContextLevel= The assessment is available at i: primary ii: secondary levels

Agency= A permanent agency, institution, or office conducts the assessments

Quality1= There is a publicly available methodology/technical document on the assessment

d. Deriving one category for Large Scale Assessment

If a country has multiple large scale assessments, the one with the highest classification achieved is retained:

$$NoInfoLSA_c = \begin{cases} 1, \text{ if } NoInfoNLSA_c = 1 \text{ and } NoInfoILSA_c = 1 \\ 0, \text{ otherwise} \end{cases}$$

$$NascentLSA_c = \begin{cases} 1, \text{ if } EstablishedLSA_c = 0 \text{ and } UnderDevLSA_c = 0 \text{ and } (NascentNLSA_c = 1) \\ 0, \text{ otherwise} \end{cases}$$

$$UnderDevLSA_c = \begin{cases} 1, \text{ if } EstablishedLSA_c = 0 \text{ and } (UnderDevNLSA_c = 1 \text{ or } UnderDevILSA_c = 1) \\ 0, \text{ otherwise} \end{cases}$$

$$EstablishedLSA_c = \begin{cases} 1, \text{ if } EstablishedNLSA_c = 1 \text{ or } EstablishedILSA_c = 1 \\ 0, \text{ otherwise} \end{cases}$$

Step 4: Assigning every country into one single category

This step gives the final status for each country.

$$NoInfo_{ck} = \begin{cases} 1, (NoInfoExam_c = 1) \text{ and } (NoInfoLSA_c = 1) \\ 0, \text{Otherwise} \end{cases}$$

$$Nascent_{ck} = \begin{cases} 1, ((NascentExam_c = 1) \text{ and } (NascentLSA_c = 1)) \text{ or} \\ ((NoInfoExam_c = 1) \text{ and } (NascentLSA_c = 1)) \text{ or} \\ ((NascentExam_c = 1) \text{ and } (NoInfoLSA_c = 1)) \text{ or} \\ ((NoInfoExam_c = 1) \text{ and } (UnderDevLSA_c = 1)) \text{ or} \\ ((UnderDevExam_c = 1) \text{ and } (NoInfoLSA_c = 1)) \text{ or} \\ ((NascentExam_c = 1) \text{ and } (UnderDevLSA_c = 1)) \text{ or} \\ ((UnderDevExam_c = 1) \text{ and } (NascentLSA_c = 1)) \\ 0, \text{Otherwise} \end{cases}$$

$$UnderDev_{ck} = \begin{cases} 1, ((UnderDevExam_c = 1) \text{ and } (UnderDevLSA_c = 1)) \text{ or} \\ ((NoInfoExam_c = 1) \text{ and } (EstablishedILSA_c = 1)) \text{ or} \\ ((EstablishedExam_c = 1) \text{ and } (NoInfoLSA_c = 1)) \text{ or} \\ ((NascentExam_c = 1) \text{ and } (EstablishedLSA_c = 1)) \text{ or} \\ ((EstablishedExam_c = 1) \text{ and } (NascentLSA_c = 1)) \text{ or} \\ ((UnderDevExam_c = 1) \text{ and } (EstablishedLSA_c = 1)) \text{ or} \\ ((EstablishedExam_c = 1) \text{ and } (UnderDevLSA_c = 1)) \\ 0, \text{Otherwise} \end{cases}$$

$$Established_{ck} = \begin{cases} 1, ((EstablishedExam_c = 1) \text{ and } (EstablishedLSA_c = 1)) \text{ or} \\ ((NoInfoExam_c = 1) \text{ and } (EstablishedNLSA_c = 1)) \\ 0, \text{Otherwise} \end{cases}$$

where,
c= country 1... country n
k= examinations and LSA combined

Step 5: Counting the number of DCPs within each of the four categories

Once the categorization of learning assessment systems as a whole is finalized by country the number of countries assigned to each category is added up for reporting purposes.

The total number of countries that fall under each of the four categories (under development $\sum UnderDev_{ck}$, nascent $\sum Nascent_{ck}$, established = $\sum Established_{ck}$, no information $\sum NoInfo_{ck}$)

Where,

$\sum NoInfo_{ck}$, $\sum Nascent_{ck}$, $\sum UnderDev_{ck}$ and $\sum Established_{ck}$ are substituted from Step 2.

Step 6: Calculate the corporate indicator value:

To calculate the proportion of DCPs meeting quality standards for LAS, $\%DCP_{MQS\ LAS}$, take the summed aggregate-level figure for countries classified as “established”, then divide that number by the total number of DCPs **C** in the sample and multiply the quotient by 100:

$$\%DCP_{MQS\ LAS} = \frac{\sum Established_{ck}}{C} * 100$$

Data limitations (if any known / anticipated):	<ol style="list-style-type: none"> 1. The results of this exercise are based on publicly available information. As such, the extent to which the results reflect the reality depends on the degree to which countries make information related to their assessment practices available online; or whether their assessment activities have already been described through initiatives focusing on learning assessments, such as the World Bank's SABER or the UIS Catalogue of Learning Assessments. 2. There may be cases where information is public, and as such it fulfills the specified criteria, but is not easily accessible via the Internet. In the future, data for this indicator will be supplemented with information collected at the country level. 3. The criteria included in this methodology for evaluating the quality of assessments are proxy measures. The presence of technical or final reports does not ensure quality; however, producing such reports and sharing them with the public suggests that some accountability measures are incorporated into the process of assessing learning achievement and an assumption is made that such measures lead to an increase in the quality. 4. This assessment cannot truly examine the extent of its alignment with the curriculum or the permanent character of the office responsible for conducting an assessment. 5. For a number of countries, the most recent assessments were not available or conducted. Hence the data gathered spans from assessments conducted between the years 2011-2015. 6. This methodology does not include classroom assessments in trying to measure the credibility of learning assessment systems. Though classroom assessments are very important, the guidelines that frame them and the quality of classroom assessment practices would be impossible to assess using publicly available information.
Interpretation	<p>A high value indicates that DCPs, by and large, have in place robust learning assessment systems to monitor progress in learning outcomes and promote evidence-based policy-making. Information on DCPs with non-established learning assessment systems, on the other hand, helps identify areas where systemic change is essential to better support countries in the strengthening of such systems.</p>

REFERENCES

- Banerji, R. & Pritchett, L.. *Schooling is Not Education! Using assessment to change the politics of non-learning*. 2013. Retrieved from: http://www.cgdev.org/doc/full_text/CGDReports/3120290/schooling-is-not-education.html
- Clarke, M.. *What matters most for student assessment systems: a framework paper (Report No. 68235)*. 2012. Retrieved from: <https://openknowledge.worldbank.org/bitstream/handle/10986/17471/682350WP00PUBL0WP10READ0web04019012.pdf;sequence=1>
- Hill, P. & Barber, M.. *Preparing for a Renaissance in Assessment*. 2014. Retrieved from: https://uk.pearson.com/content/dam/region-core/uk/pearson-uk/documents/educator/primary/preparing_for_a_renaissance_in_assessment_and_summary_text_december_2014.pdf
- Learning Metrics Task Force. *Toward Universal Learning- Implementing Assessment to Improve Learning*. 2014. Retrieved from: <http://www.uis.unesco.org/Education/Documents/towards-universal-learning-a-global-framework-for-measuring-learning-metrics-task-force-education-2013-en.pdf>
- Education Policy Data Center, FHI 360. *National Learning Assessment Mapping Project*. 2015.
- Wagner, D. *Smaller, Quicker, Cheaper: Improving Learning Assessments for Developing Countries*. 2011. Retrieved from: <http://unesdoc.unesco.org/images/0021/002136/213663e.pdf>

ANNEXES

Annex 1- Data Collection tool

Data collection tool utilized for collecting the data, if any:

“Quality standards for learning assessments systems” questionnaire

Annex 2- Standard Operating Procedure

Annex 3- Additional Analysis

Each category is analyzed based on the total number of countries that fall into each category. A distribution analysis is then conducted to assess which countries perform are underperforming in comparison to their peers so as to provide more capacity and technical assistance to these countries.