United Nations
Educational, Scientific and
Cultural Organization

UNESCO INSTITUTE FOR STATISTICS

GLOBAL ALLIANCE TO MONITOR LEARNING

SUSTAINABLE DEVELOPMENT GOALS

Minimum proficiency levels (MPLs): outcomes of the consensus building meeting

Background papers

GAML Fifth Meeting
17-18 October 2018
Hamburg, Germany

This package has been prepared as background documents for the Consensus Building Meeting on Proficiency Levels that took place on 10-11 September 2018 in Paris.

Contents

Paper 1 - Mathematics -Methodology for ordering performance level descriptors

This paper is presented to describe the methodology utilized for analysing, comparing, simplifying, and ordering the performance level descriptors for various national and multinational assessments against the UIS Proficiency Scale in mathematics

Background

Indicator 4.1.1

The UNESCO Institute for Statistics' (UIS) goal as a custodian agency for reporting against the Sustainable Development Goals in Education (SDG4) is to develop standards, methodology and guidelines to enable countries in the production of data for the reporting of indicators Indicator 4.1.1. requires member countries to report on the "proportion of children and young people….to achieve at least a minimum proficiency level in reading and mathematics". In order to define the minimum proficiency for report indicator 4.1.1, the UIS has developed the Global Framework for Mathematics, organized and compiled cross-national assessment performance level descriptors, with the goal of building consensus on the number of performance levels, definition of policy and performance descriptors, as well as of the minimum proficiency levels for each education level.

List of Assessments

The assessments for which PLD's were analyzed for this project are shown in Table 1, The assessments were grouped into three grade-level bands, or measurement points: 2-3; 4-6; and 8-9.

Table 1. Assessment Information.

Assessment Name	Assessment Type	LSA	Year Administered
ASER	National Citizen-Led	Grades 2-3	2017
EGMA	National	Grades 2-3	Not provided
PASEC	Regional	Grades 2-3	2014
TERCE	Regional	Grades 2-3	2014
UNICEF MICS6	Household Survey	Grades 2-3	Not provided
Uwezo	National Citizen-Led	Grades 2-3	Not provided
PASEC	Regional	Grades 4-6	2014
PILNA	Regional	Grades 4-6	2015
SACMEQ	Regional	Grades 4-6	2007
TERCE	Regional	Grades 4-6	2014
TIMSS	Cross-national	Grades 4-6	2015
PISA	Cross-national	Grades 8-9	2015
PISA-D	Cross-national	Grades 8-9	Not provided
TIMSS	Cross-national	Grades 8-9	2015

Performance Level Descriptors

Definition

Each assessment in Table 1 has a number of performance level descriptors (PLD's) associated with it. These PLD's delineate one or more mathematical skills and/or processes that are associated with test takers who achieve that performance level. The number of PLD's varies by assessment, as does the format in which the PLD's are written. Examples of mathematical skills include counting, adding fractions, solving equations; examples of mathematical processes include employing basic formulas, interpreting problem situations, and communicating reasoning.

Analysis, comparison, and ordering

The primary, if not sole, criterion for analysing PLD's is the *cognitive demand* required by the mathematical skills and/or processes contained in each PLD. This is complicated by the fact that most, if not all, PLD's contain multiple skills and processes. Thus, comparing PLD's becomes a matter of determining and comparing the *overall* cognitive demand of each PLD. This requires a high level of careful analysis, and is as much art as science. Successively comparing PLD's against each other eventually resulted in a list of the PLD's within each measurement point, arranged from lowest to highest overall cognitive demand. As an additional point of information, each PLD was given a one-sentence summary, which may facilitate easier comparison for future work.

Proficiency Scale

Once the list of PLD's for each measurement point was completed, it was then necessary, and possible, to create the overall Proficiency Scale for mathematics. This was begun by placing *all* the PLD's from *all* three measurement points into a single list, from the lowest of grades 2-3 to the highest of grades 8-9. However, it could not be assumed that the highest-level PLD of one measurement point had a lower cognitive demand than the lowest-level PLD of the next-highest measurement point. The next step was therefore to compare the high-level PLD's of grades 2-3 against the low-level PLD's of grades 4-6, utilising the same process of comparing the overall cognitive demand of the PLD's, and re-arranging PLD's as appropriate. This was then repeated with the PLD's at the border of grades 4-6 and grades 8-9. This resulted in a list of *all* PLD's across all three measurement points.

Ordering within measurement points

The final step after creating the Proficiency Scale was to identify which PLD's contained grade-level appropriate (GLA) skills and processes for each measurement point. For this step, cognitive demand was *not* a criterion, as each measurement point contains a range of skills from low to high cognitive demand. The Proficiency Scale includes a number of PLD's that did not contain GLA skills or processes even at the lowest measurement point. It also included many PLD's that were GLA at more than one measurement point.

Once the Proficiency Scale was complete, it was then possible to set the performance levels at each measurement point, using the list of GLA PLD's. Each measurement point used the same four performance levels—Below Basic; Basic; Proficient; and Advanced. As with the first step in the process, determining where to set each performance level required a good deal of careful analysis, especially since the skills and processes taught at each grade can vary, in some cases

widely, from nation to nation. Finally, at each measurement point, the lowest PLD in the Proficient performance level was marked as the dividing line between proficient less than proficient test takers. See **Figure 1** for an excerpt of the Proficiency Scale.

Figure 1. UIS Proficiency Scale (excerpt).

Assessment Name	Assessment Level	Domain	Performance Level	Descriptor	One line descriptor	Order	GR 2-3	GR 4-6	GR 8-9
SACMEQ 2007 (grade 6)	Regional	Mathematics	**Level 1**	Pre Numeracy Applies single step addition or subtraction operations. Recognizes simple shapes. Matches numbers and pictures. Count in whole numbers.	Students at this level can perform only the simplest of computations and can recognize simple shapes.	42	X		
PILNA 2015 (grades 4/6)	Regional	Mathematics	**Level 2** *425-449*	• Write a three-digit number not involving zero in words and in numerals, and write a three-digit number involving zero in words only. • Compare prices of items and calculate the total cost of two items. • Subtract a two-digit number from a two- or three-digit number without regrouping and solve simple word problems involving addition. • Identify hands of a clock and know the relation of days and weeks.	Students at this level can write, compare and compute with whole numbers.	43	X		
TERCE 2014 (grade 3)	Regional	Mathematics	**Level 4** *843+*	Students can: • Solve more complex problems in the area of natural numbers. • Solve problems involving comparison and conversion of measures • Solve more complex problems involving elements of geometric figures or flat representation of geometric shapes.	Students at this level are able to utilize reasoning skills to solve complex mathematical and real-world problems.	44	X	X	

Mapping

Once the PLD's were placed in order, the final task was to create a graphical display, or mapping, of each assessment's PLD's against the performance levels at each measurement point, as well as an overall mapping of all assessments. This overall mapping is not compared against performance levels, but is mapped against the grade-level progression, in order to show where the individual PLD's for each assessment lie in comparison to each other. It should be noted that those PLD's that were considered to be below the minimum level for the grades 2-3 measurement point were *not* included in the mapping for that measurement point, or for the overall mapping.

As is typical of assessments, each performance level represents a range of abilities on the part of test takers. This range is usually represented by scale scores, which are provided for most of the assessments in this project. However, each assessment uses a different scale, so a direct comparison between scale scores is not possible. Because the performance levels were set without the benefit of scale scores, a decision was made to map the space for the performance levels proportionally to the ordered placement of the PLD's at each measurement point. For example, at grades 2-3, there are 3 spaces separating TERCE Level 1 and Level 2. Thus, the TERCE Level 2 bar takes up 3 columns in the mapping.

The final step in creating the overall mapping was to "fill in the blanks" that existed between performance levels within an assessment when the mappings for all three measurement points were placed onto the overall mapping. For instance, for PASEC grade 6, the bar for "Below Level 1" goes part way across grades 2-3, while "Level 1" begins in grades 4-6. In order to "fill in the gap" on the overall mapping, the "Level 1" bar was extended *backwards* until it "met" the "Below Level 1" bar. This was done as a way of indicating that test takers can, and most likely will, achieve different levels of achievement across the grade-level continuum.

Policy Level Descriptors

Previously, policy level descriptors in the area of mathematics were developed to characterize (in general terms) the difference in ability between mathematically proficient test takers and those who achieve at a level below proficiency. These policy level descriptors reflect the dividing line between proficient and non-proficient test takers, even though they do not delineate between the two sub-categories at each level: Below Basic vs Basic, and Proficient vs Advanced. The policy level descriptors are an exceedingly useful and important tool that can be used to validate that the content described at each measurement point is an accurate reflection of the mathematical skills and processes for which students around the world should be expected to demonstrate a certain degree of mastery.

Figure 2. Policy Level Descriptors for Mathematics.

Performance Level Policy Descriptors	
Proficient/ Above Proficiency	Students at this level possess a basic, or better, level of mathematical knowledge. They also demonstrate a basic, or better, level of competency with mathematical skills and abilities. These includes the recall of mathematical facts, formulas, and algorithms, the ability to solve application problems, and varying levels of aptitude in using problem-solving strategies and communicating mathematically.
Below Proficiency	Students at this level possess a limited level of mathematical knowledge and demonstrate a lack of competency with most mathematical skills and abilities. They tend to struggle with all but the most routine and straightforward aspects of mathematics.

Paper 2: Mathematics -Methodology for PLD compilation and cross-functional alignment

This paper is presented to explain the methodology utilized in the development of two documents:

1) an aggregated compilation of the performance level descriptors (PLD's) of various international assessment instruments, and
2) an alignment of the international assessment PLD's to the UNESCO Global Framework for School Mathematics (GF).

Methodology for compiling PLD's

The overall goal for the first document was to create two sets of compiled PLD's around a single cut point, thereby creating two categories—Proficient/Above Proficiency and Below Proficiency (the former will be shortened to "Above Proficiency" throughout the rest of this paper). Table 1 shows the descriptors for these two categories.

Table 1. General definitions of performance levels.

Performance Level Policy Descriptors	
Proficient/ Above Proficiency	Students at this level possess a satisfactory, or better, level of mathematical knowledge. They also demonstrate a satisfactory, or better, level of competency with mathematical skills and abilities. These includes the recall of mathematical facts, formulas, and algorithms, the ability to solve application problems, and varying levels of aptitude in using problem-solving strategies and communicating mathematically.
Below Proficiency	Students at this level possess a limited level of mathematical knowledge and demonstrate a lack of competency with most mathematical skills and abilities. They tend to struggle with all but the most routine and straightforward aspects of mathematics.

These policy descriptors are applied at three measurement points: grade levels 2-3; grade levels 4-6; and grade levels 8-9. The first step in determining the cut point was to determine a common cut point among the performance levels (PL's) of the various assessments, each of which has anywhere from 3 to 8 PL's. This information is shown in Table 5 (not included here), provided by UNESCO personnel. In this table, the PL's above the cut point are highlighted in blue—e.g., Levels 2-6 of the grade PISA assessment.

Despite the varying number of PL's in the different assessments, the PLD's themselves are all of a very similar format—each being comprised of several skill descriptors, which are typically presented as bullet points, or as individual sentences. (The term "skill descriptor" may be a bit misleading, as many, if not most, of them actually contain *more* than one skill; e.g., "model and solve equations".) The next step in developing the aggregate PLD's was thus to analyze all of the individual skill descriptors for each category in a measurement point, such as Above Proficiency for grades 8-9. This analysis was focused on the *cognitive process* required for each skill, as

described by the literal text of the individual skill descriptors. This analysis served two purposes: first, it identified descriptors that described, to varying degrees, the same skill—be it between PLD's of the same assessment, or between assessments, or both. Second, it identified closely related skills that could be combined into a single skill descriptor. Finally, the results of this analysis were used to create the aggregate PLD's. This basically involved rewording the original skill descriptors to the extent possible and/or necessary, based on the analysis described above, the removal of redundant language, and the combining of skill descriptors where appropriate. Care was taken when combining skill descriptors to lessen the overall word count from the original assessments' PLD's without sacrificing clarity or the mathematical meaning of the original text. As an example, in the Above Proficiency category for grades 2-3, there were 21 individual skill descriptors containing 326 words. This was shortened during the compilation process to a slightly less unwieldy 18 skill descriptors, comprised of a much more readable 187 words.

The resulting PLD's differ to varying degrees in number of words and skill descriptors, and in coverage of the various domains of the Global Framework. This can be attributed to the different number of assessments in various measurement points; the different number of PL's in each assessment; and, of course, the differences in the PLD's themselves. Even the way in which an individual assessment's PLD's were written was a contributing factor in some cases. It should also be noted that very little latitude was given when analyzing the mathematical language of the assessment PLD's for identification of skills and cognitive processes, unless analysis of other PLD's warranted otherwise. That is, mathematical language was taken *literally* whenever possible; for example, skill descriptors involving "algebraic expressions" were taken to mean polynomials and such, rather than interpreting the term more widely to include inequalities and equations (in American judicial terms, this practice is known as "strict constructionism"). This approach was utilized as a matter of overall alignment philosophy and is critical in creating a product that matches the original assessment PLD's as closely as possible. Only in cases of ambiguous language or meaning in the assessment PLD's was any interpretation allowed, and in those cases, the goal was always to divine the original intent.

Methodology for aligning PLD's to the Global Framework

The alignment of the assessment PLD's to the Global Framework utilized the previously described analysis of PLD language to identify the cognitive process described in each individual skill descriptor. However, before this could take place, it was first necessary to determine to which level of granularity of the GF the PLD's would be aligned. The sub-construct level of the GF is closest in format to the PLD's in its descriptions of individual mathematical skills; it thus made sense to choose this level for alignment to the PLD's. It was also necessary to perform an analysis of the GF sub-constructs to determine the necessary cognitive process for the skills described in each sub-construct (a painstaking process, as some sub-constructs contain descriptions of over 20 skills).

Once the analyses of the two documents were complete, it was then possible to perform the alignment. An alignment between a GF sub-construct and a PLD skill descriptor was determined to exist when the *same cognitive process* is present in both areas. In the alignment document, the text of the skill descriptor is entirely or partially bold-faced; the bold-faced text indicates the part (or entirety) of the descriptor that reflects the cognitive process that determined the alignment. In some cases, more than one cognitive process alignment exists between a sub-construct and a

skill descriptor; other than the aforementioned bold-faced text, no indication of multiple alignments was deemed necessary, and thus none is indicated in the alignment spreadsheet.

Because many assessment PLD's describe the same or closely related skills, many of the GF sub-constructs were aligned to multiple skill descriptors; this is also due, in many cases, to the wide range of skills described by some sub-constructs. This even led to instances where the same skill is present in both the Above Proficiency and Below Proficiency categories at a measurement point. Conversely, there were a number of sub-constructs that had no alignment to any skill descriptors. Finally, there were also a handful or skill descriptors that did not align to any sub-constructs; these skill descriptors are listed at the bottom of the alignment spreadsheet.

In terms of individual assessments, Tables 2 and 3 display the alignment of content coverage based on the PLD's from each assessment. Table 2 displays each assessment's coverage of each domain in the GF, using one of three categories—Minimal, Moderate, or Extensive. These categories reflect the number of alignments between each assessment's PLD's and the sub-constructs in each domain—a rating of Minimal indicates from 1-3 alignments; Moderate, 4-6 alignments; Extensive, more than 6 alignments. Table 3 displays each assessment's coverage of the GF sub-domains. However, instead of categorizing the number of alignments at each sub-domain, as in Table 2, Table 3 merely indicates the presence (or absence) of one or more alignments.

Table 2. Analysis of the coverage of the Global Framework for School Mathematics domains, based on the Performance Level Descriptors of international assessments.

		GLOBAL FRAMEWORK DOMAINS					
TEST	GRADE	MATH PROFICIENCY	NUMBER KNOWLEDGE	MEASUREMENT	STATISTICS	GEOMETRY	ALGEBRA
EGMA	N/A		Extensive			Minimal	Minimal
ASER	N/A		Moderate				
UNICEF MICS6	N/A		Moderate				Minimal
Uwezo	N/A		Extensive				
PASEC	2		Extensive			Minimal	Minimal
SERCE	3		Moderate	Extensive	Minimal	Extensive	Extensive
TERCE	3		Extensive	Extensive	Minimal	Extensive	Minimal
TIMSS	4	Minimal	Extensive	Extensive	Extensive	Extensive	Moderate
PILNA	4/6*		Extensive	Extensive	Minimal		Moderate
PASEC	6	Moderate	Extensive	Extensive	Minimal	Minimal	Minimal
SACMEQ	6	Extensive	Extensive	Extensive	Minimal		Moderate
SERCE	6		Extensive	Extensive	Minimal	Extensive	Moderate
TERCE	6		Extensive	Extensive	Moderate	Moderate	Extensive
PISA	8	Extensive	Extensive				Minimal
PISA-D	NA	Extensive	Extensive				Minimal
TIMSS	8	Moderate	Extensive		Extensive	Extensive	Extensive

*The Performance Level Descriptors for PILNA overlap between grades 4 and 6.

Table 3. Analysis of the coverage of the Global Framework for School Mathematics sub-domains, based on the Performance Level Descriptors of international assessments.

		GLOBAL FRAMEWORK SUB-DOMAINS															
TEST	**GRADE**	**1.1**	**1.2**	**1.3**	**2.1**	**2.2**	**3.1**	**3.2**	**4.1**	**4.2**	**5.1**	**5.2**	**5.3**	**6.1**	**6.2**	**6.3**	**6.4**	**6.5**
EGMA	N/A				X	X												
ASER	N/A					X												
UNICEF MICS6	N/A					X									X			
Uwezo	N/A				X	X												
PASEC	2				X	X					X	X			X			
SERCE	3					X		X	X		X	X	X	X	X	X		X
TERCE	3					X		X	X		X		X		X			
TIMSS	4		X	X		X		X	X		X		X			X		X
PILNA	4/6*					X		X	X						X			X
PASEC	6	X				X		X	X		X	X	X		X			
SACMEQ	6	X	X		X	X		X	X							X		X
SERCE	6					X		X	X		X				X			X
TERCE	6					X		X	X		X		X		X			X
PISA	8	X	X	X		X												X
PISA-D	N/A	X	X	X		X												X
TIMSS	8	X	X			X			X	X	X		X		X	X		X

*The Performance Level Descriptors for PILNA overlap between grades 4 and 6.

Paper 3: Reading -Compilation of performance level descriptors across regional and international assessments

This paper aims at answering the following questions:

1. How do performance level descriptors from regional and international reading assessments compare to each other?
2. Which should be the MPLexpected in reading for Grades 2 & 3, end of Primary education and end of low Secondary education?
3. Which should be the MPLexpected in reading overall?

Based on the analysis of regional and international assessments of reading described on a previous paper, this document aims to facilitate the process of setting common expectations between different cross-national assessments to allow for international comparison.

In order to answer the first question, this paper shows the process by which the Performance level descriptors (PLDs) of all of the regional and international assessments on reading are analyzed, ordered according to difficulty and grouped into four performance categories. This is done for each educational level considered in the 4.1.1 indicator of SDG 4, which are Grades 2 & 3 (4.1.1a), the end of Primary education (4.1.1b) and at the end of Low Secondary education (4.1.1c).

However, consideration has to be given to the fact that this analysis is based on the PLDs only; therefore, the assessments' aims as well as the tasks used by them have not been considered. Even though some of the PLDs make explicit the type of text they use, most of them do not. Thus, the ordering process was done regarding the cognitive demand implied by the processes mentioned in the PLDs. This may lead to assumptions regarding difficulty that are not congruent if a task-based analysis is performed, altering consequently the order of the PLDs. For example, even though from a cognitive perspective making inferences is more difficult than retrieving explicit information, making an inference from a short narrative text is likely to be easier than retrieving explicit information from a long technical informative text. Therefore, an analysis of the PLDs together with task type information provided by the assessment frameworks may lead to a more precise ordering.

Furthermore, to answer questions 2 and 3, a Minimum proficiency level (MPL) was set for each educational level and for reading acquisition in general with accompanying policy descriptors.

Characteristics of the regional international assessments

Table 1 shows the cross-national assessments considered for this paper. Most of these tests are designed to evaluate formal learning, as is the case of reading. However, both ASER and UNICEF MICS 6 are broader questionnaires that aim at obtaining other development indicators at a personal, family, and environmental level, which include a section on reading that is the one considered in this analysis.

Table 1. Characteristics of the assessments

Name	Abbreviation	Grade/Age	Corresponding SDG 4 indicator	Minimum proficiency level	Observations
Annual Status of Education Report	ASER	6 to 14 year-olds	4.1.1.a;	Standard 2 (story)	Part of a household questionnaire in which the assessment is individual.
UNICEF Multiple Indicator Cluster Service	UNICEF MICS 6	5 to 17 year-olds	4.1.1.a;	Foundational Reading Skills	Part of a household questionnaire in which the assessment is individual.
UWEZO Annual Learning Assessment	UWEZO	6 to 16 year-olds	4.1.1.a;	Standard 2	Part of a household questionnaire in which the assessment is individual.
Early Grade Reading Assessment	EGRA	Grades 1 to 3.	4.1.1.a	Not specified	Individual assessment
Third Regional Comparative and Exploratory Study	TERCE	Grades 3 & 6	4.1.1.a; 4.1.1.b	Level 2	School-based assessment
Pacific Islands Literacy and Numeracy Assessment	PILNA	Grades 4 & 6	4.1.1.b	Level 4 (grade 4) and Level 5 (grade 6).	School-based assessment
Progress in International Reading Literacy Study	PIRLS	Grade 4	4.1.1.b	Low international Benchmark (second level)	School-based assessment
The Analysis Programme of the CONFEMEN Education Systems	PASEC	Grades 2 & 6	4.1.1.a; 4.1.1.b	Level 3	School-based assessment. Partly individual assessment

Name	Abbreviation	Grade/Age	Corresponding SDG 4 indicator	Minimum proficiency level	Observations
Southern and Eastern Africa Consortium for Monitoring Educational Quality	SACMEQ	Grade 6	4.1.1.b	Level 3	School-based assessment
Programme for International Student Assessment	PISA and PISA-D	15 year-olds	4.1.1.c	Level 2	School-based assessment

Characteristics of the regional international assessments

The initial step taken in this process to answer the first question was to develop a Proficiency Scale (PS) on reading. In this regard, all of the PLDs across the ten assessments analyzed were transformed into one-line descriptors by highlighting its main characteristics and those that differentiated them from the previous level.

After this, all of the descriptors were ordered according to their difficulty independently from the educational level they were designed for. This produced a 73 level PS that considers all of the PLDs provided by the ten assessments. It is important to note that the below level 1 descriptor from PASEC as well as the Level 0 descriptor from PILNA were not considered as there is no specific information regarding what the student can or cannot do in those levels.

An interesting finding that arises from the development of the PS is the incongruence between the expectations set by different regional and international assessments as well as the overlapping of PLDs designed for different educational levels.

Finally, in order to answer the third question, an overall MPL was set for reading in general. This was marked at the 50th level on the PS that corresponds to TERCE's Level 2 performance descriptor for Grade 3 which is summarized as: "Students understand the global sense of the text by distinguishing its central topic and making inferences regarding non evident information". If we analyze it from the Global Framework for Reading perspective, it assumes mastery of the decoding sub domain as well as explicitly includes the retrieve and interpret constructs from the reading comprehension sub domain. Even though the other constructs that correspond to the reading comprehension subdomain (reflect, metacognition and motivation and disposition) are desirable, these are not necessary for most of the reading tasks people are faced with in everyday life.

Figure 1 shows the PS and the overall MPL. **Figure 2** shows the performance descriptors that are above the MPL.

Figure 1.

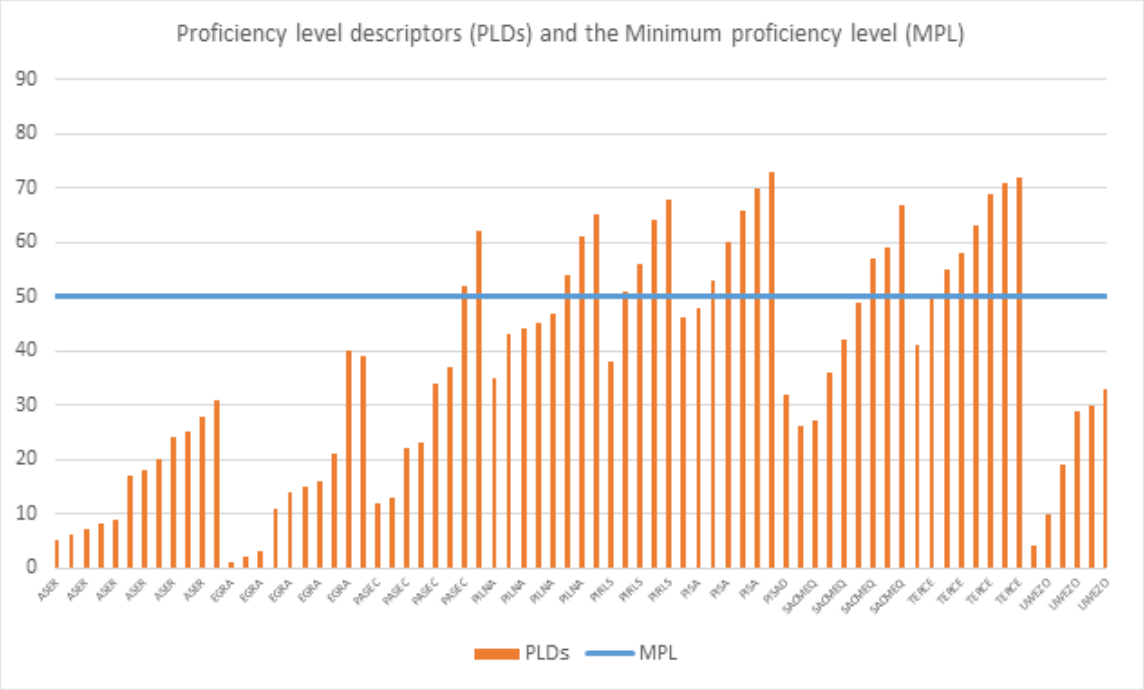
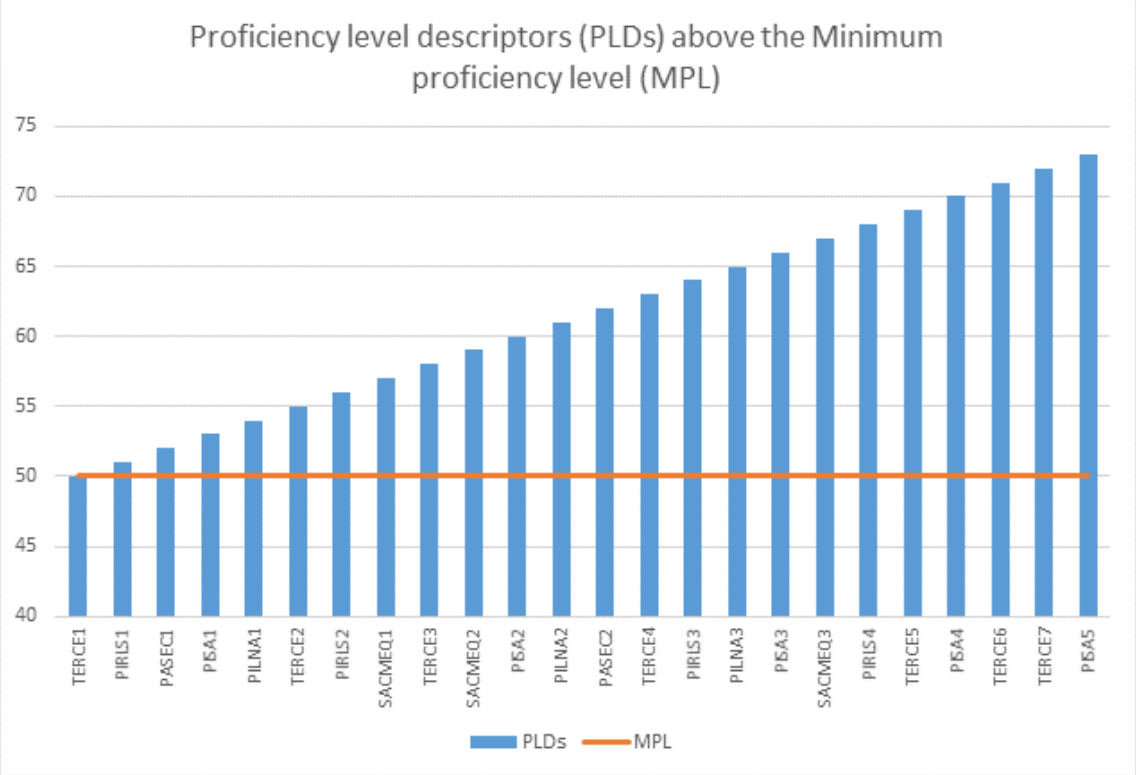


Figure 2.



Characteristics of the regional international assessments

The 73 levels of the PS were divided into the three educational levels considering their levels of difficulty as well as the acquisition of skills these entailed. This constitutes the reference scales.

For all of the educational levels the descriptors included in the reference scale spanned from below basic level expected for that grade to advanced knowledge. Therefore, numerous performance descriptors overlap between educational levels.

Subsequently, the performance descriptors that compose each reference scale were divided into four categories according to difficulty. These categories are below basic, basic, proficient and advanced.

The below basic category is constructed based on descriptors that are expected to have already been achieved by the start of the educational level. The basic category, on the other hand, is composed by the performance descriptors that reflect the minimum skills to be acquired during that educational level. The highest descriptor of this category will constitute the MPL expected for that educational level. Moreover, the proficient category entails skills that, though being over the minimum expected, may be developed during the grade by an important percentage of students. Finally, the advanced category was developed in order to be able to consider those students that show very good reading skills.

The next three sections will answer the second question by describing how the PLDs from different assessments map into the reference scale developed for each educational level. A comparison between the MPLs set by each regional and international assessment and the MPL established according to the reference scale will be drawn.

Grade 2 & 3 (4.1.1.a.)

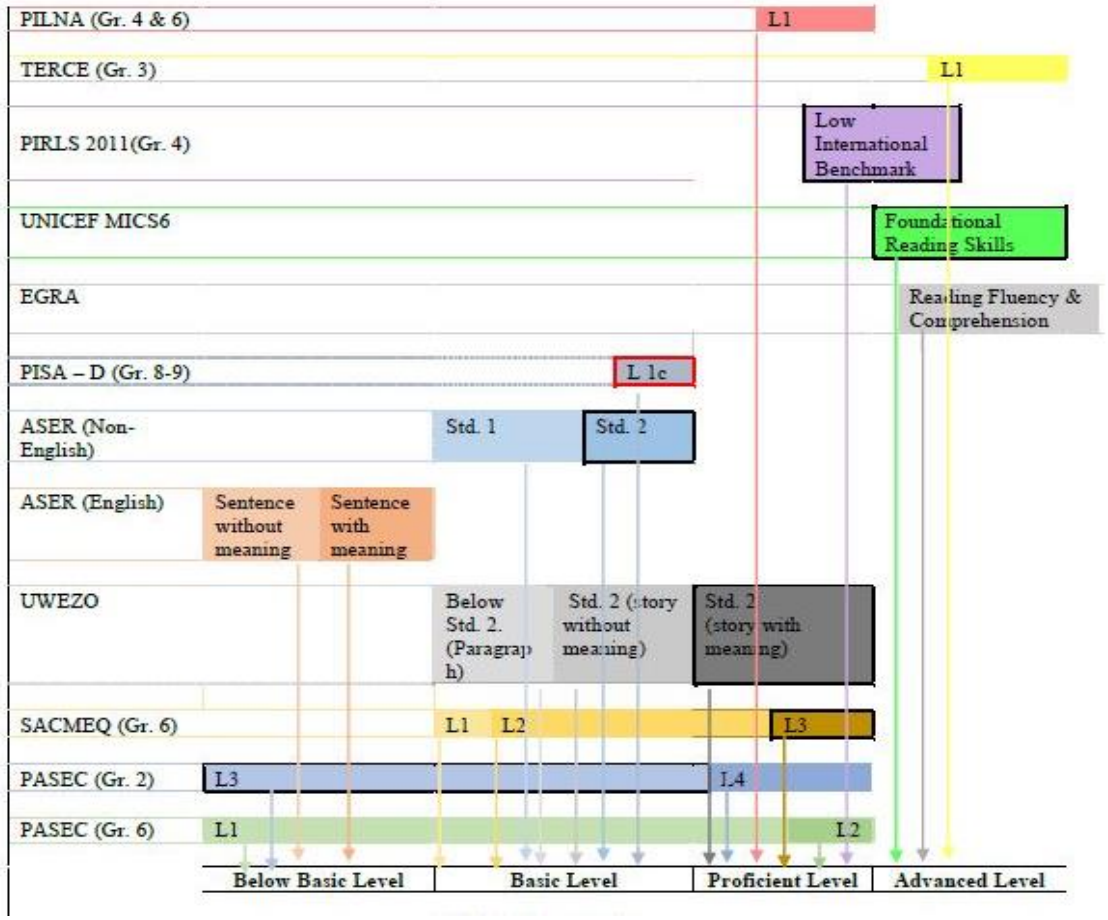
The reference scale for grades 2 & 3 is constituted by 20 PLDs that go from level 22 to 41 from the PS. Levels 22-25 belong to the below basic category, 26-32 to the basic category, 33-38 to the proficient category and 39-41 to the advanced category.

The MPL set for grades 2 & 3 is level 32 from the PS which corresponds to Level 1c from PISA for Development (PISA-D) which is summarized as “students understand the meaning of sentences and very short simple passages with familiar contexts”. This is considered the minimum to be expected for this educational level because it implies having achieved mastery regarding precision in decoding, but not necessarily fluency in this sub domain. Moreover, it builds on students’ linguistic knowledge by considering familiar contexts and assumes retrieving of simple explicit information.

Figure 3 shows how the different PLDs from the regional and international assessments map into the reference scale for this educational level. The assessment levels that are highlighted by black borders are the established as minimum proficiency by each assessment. While the performance level highlighted with red borders corresponds to the one explained in the previous paragraph.

As can be concluded from the figure above, ASER’s (Non English) and PASEC’s (Grade 2) MPLs are easier than the MPL set in the reference scale. Moreover, TERCE’s Level 1 is significantly more difficult than the MPL established for this educational level, being considered in the advanced category.

Figure 3.



Finally, there is a surprising overlapping between assessments designed for different educational levels, being both SACMEQ's (grade 6) and PIRLS 2011's (Grade 4) MPL considered as proficient for grades 2 & 3 not far from this educational level's MPL. Moreover, considering ASER, UWEZO and UNICEF MICS 6 as the assessments with a broader application spectrum that cover up to the third educational level, it is interesting that their minimum MPLs correspond to the basic, proficient and advanced categories for grades 2 & 3 respectively.

End of primary education (4.1.1.b.)

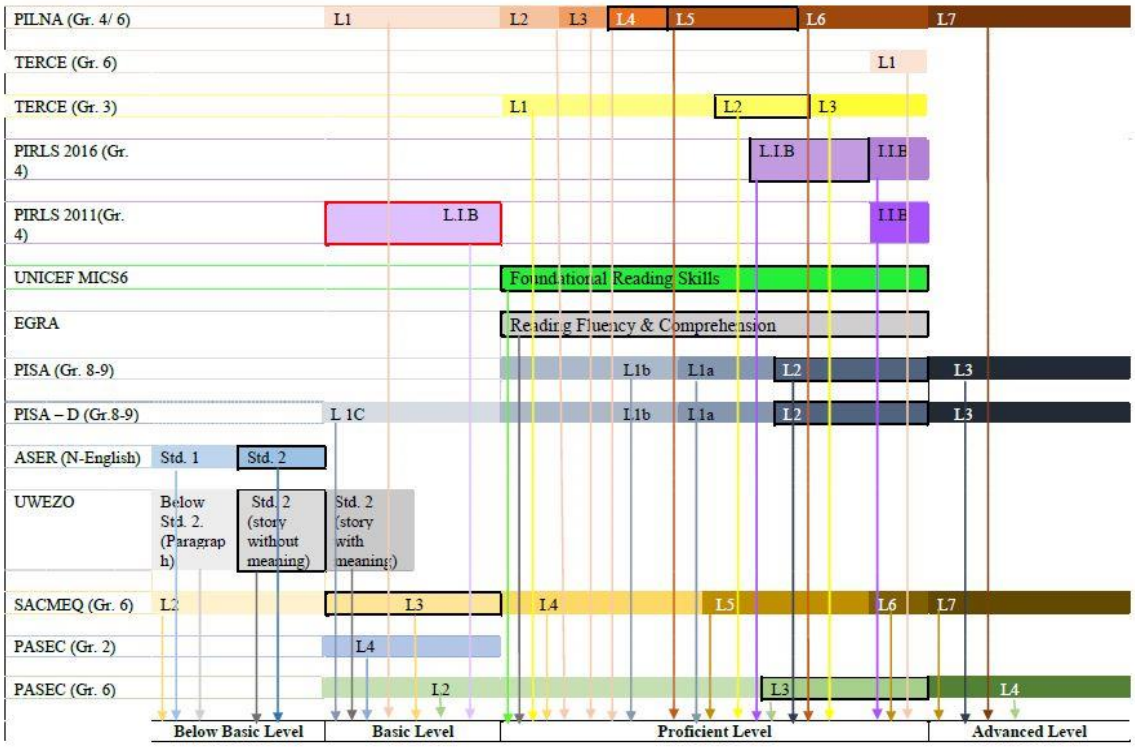
The reference scale for grades 4 & 6 is constituted by 36 PLDs that go from level 27 to 62 from the PS. Levels 27-31 belong to the below basic category, 32-38 to the basic category, 39-58 to the proficient category and 59-62 to the advanced category.

The MPL set for the End of Primary is level 38 from the PS that corresponds to Low International Benchmark from PIRLS 2011 which is summarized as "students identify and retrieve explicit information from informational and literary texts". This is considered to be the minimum to be expected for this educational level because it implies having achieved mastery regarding decoding as well as having developed at least the possibility of identifying different types of texts and retrieving explicit information from them.

Figure 4 shows how the different PLDs from the regional and international assessments map into the reference scale for this educational level. The assessment levels that are highlighted

by black borders are the established as minimum proficiency by each assessment. While the performance level highlighted with red borders corresponds to the one explained in the previous paragraph.

Figure 4.



As can be concluded from the figure above, ASER's (Non English) and UWEZO's MPLs are easier than the MPL set in the reference scale for the end of Primary Education. Even though, the same happens with SACMEQ's, this is closer to the MPL set for this educational level. Moreover, PILNA's MPLs both for grades 4 & 6 are more difficult than the one that has been set, the same happens with PASEC's for grade 6. A very interesting difference is the one that exists between PIRLS 2011 and PIRLS 2016 Low International Benchmark, being the latter significantly more difficult than the former.

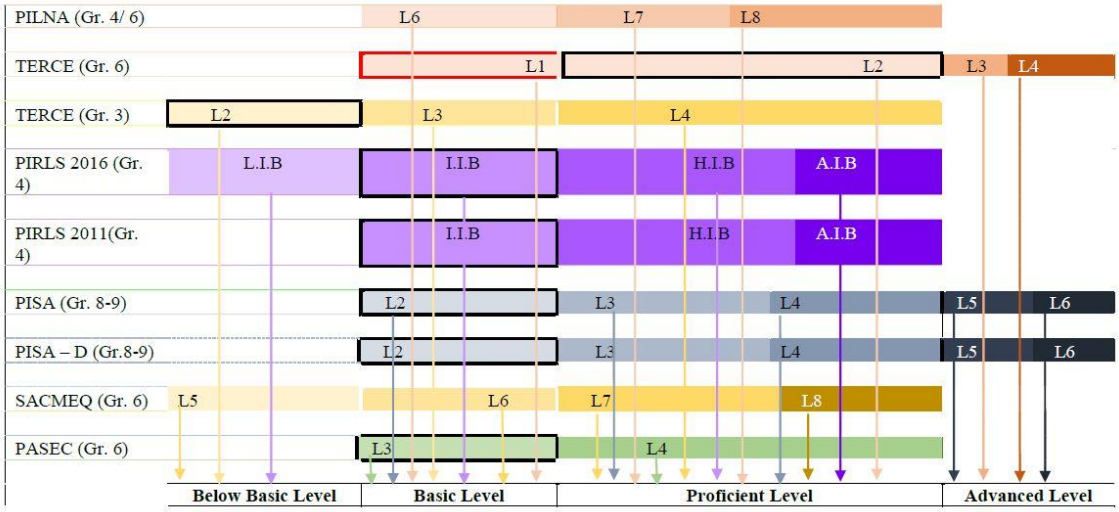
Finally, there is a surprising overlapping between assessments designed for different educational levels. The difference between the minimum levels of proficiency expected by TERCE (grade 3), PASEC's (Grade 6) and PISA (Grades 8 & 9) is surprisingly small considering the grade variation.

End of lower secondary education (4.1.1.c.)

The MPL set for the End of Low Secondary is level 58 from the PS which corresponds to Level 1 from TERCE (Grade 6) which is summarized as "students make causal relations among information from a text and can identify the issuer of a text." This is considered to be the minimum to be expected for this educational level because it implies having achieved mastery regarding decoding as well as being able to retrieve explicit information, interpret the information given by relating it to previous knowledge and reflect upon information from the text as well as its author.

Figure 5 shows how the different PLDs from the regional and international assessments map into the reference scale for this educational level. The assessment levels that are highlighted by black borders are the established as minimum proficiency by each assessment. While the performance level highlighted with red borders corresponds to the one explained in the previous paragraph.

Figure 5.



As can be seen from the figure above, ASER, UWEZO and UNICEF MICS 6 do not appear, as the PLDs used by these assessments are significantly easier than what is expected for this educational level, even though the age range of application corresponds. Furthermore, it is important to note that PISA's MPL is also easier than the one established for this educational level.

Finally, there is evident overlapping between MPLs from different assessments, especially when considering all of the performance levels that correspond to the basic category, in which we can find the minimum proficiency levels expected by PIRLS for grade 4, PASEC for grade 6 and PISA for grade 9. Moreover, there is great overlapping in the advanced category between PISA's highest two levels and TERCE's (grade 6) highest two levels of performance, which is unexpected as there are two grades in between.

After analyzing the three educational levels separately, a summary of the MPLs set for each of them, as well as the overall one will be presented together with the policy descriptor for minimum proficiency.

Analyzing the minimum proficiency levels in the light of policy descriptors

In a previous paper, the process of developing policy descriptors was explained. From that process, a policy descriptor for achieving minimum proficiency in reading was created. That descriptor stated, "Students have developed the required competences for the described reading level. They have acquired the knowledge and skills necessary to decode written words, identify relevant information from written texts, understand their meaning and make inferences from their knowledge".
This section will look at the three MPLs in the light of the policy descriptor.

In the case of Grades 2 & 3 the MPL is "students understand the meaning of sentences and very short simple passages with familiar contexts". From this perspective, the required competences to be developed in order to achieve this level are precision in decoding individual words and sentences as well as retrieving explicit information from very short passages.

For the end of Primary Education, the MPL set was "students identify and retrieve explicit information from informational and literary texts". In this regard, the competences necessary to achieve this level are precision and certain degree of fluency in decoding, as well as, the identification of different types of texts and retrieving explicit information from them.

Finally, for the end of Low Secondary Education the MPL established is "students make causal relations among information from a text and can identify the issuer of a text". Even though, not explicitly stated in the descriptor, this level implies having developed mastery in decoding regarding both precision and fluency, having achieved a literal comprehension of different types of texts, being able to interpret implicit information from different parts of the text as well as reflect upon the source of the text and its author.

Conclusions and recommendations

Overall, analyzing the MPLs for Reading at three educational cut-points allows for a better understanding on how the process of reading acquisition is expected to develop through formal schooling.

Even though, language and cultural differences may influence the rate of development, generating differences between countries at certain stages, it is believed that the MPL descriptors are specific enough to be measurable and at the same time broad enough to be adjustable to different languages and cultures.

In order to increase international comparability between assessments, agreement has to be reached related to the processes and skills being assessed and the level of development to be expected at each educational level.

In this sense, an option would be to create for each educational level a MPL, but at the same time, to separate that level into processes or skills, being able to assess student's achievements in those separately. In this model, different countries may achieve the MPL at a given educational stage in some processes and skills but not in all of them. This is similar to the model proposed by ACER's Learning Progression Explorer. Therefore, this would take into account country variability, while at the same time increasing comparison potential. Considering the constructs from the Global Framework for Reading in order to establish these processes and skills may prove to be useful.

Furthermore, another way of increasing comparability between regional and international assessments would be to make explicit in the PLDs some information regarding the tasks used. The main task characteristics that may affect PLD difficulty and therefore comparability would be:

a) Text type: continuous or discontinuous; narrative, descriptive, informational, etc.
b) Text length: overall text length as well as how sparse in the text is the information needed to perform the task.

c) Text topic or meaning: is the topic of the text known to students, are they expected to have previous knowledge about it, would its meaning be clear to them.

d) Vocabulary: use of familiar or non-familiar words, use of technical vocabulary.

e) Different sources of information: does the task involve considering more than one source of information, for example: textual and paratextual information (images, tables, graphs, figures), more than one text.

A description of task characteristics, together with the processes being assessed, could aid in evaluating the overall cognitive demand and difficulty of any given performance level descriptor.

Paper 4: Reading - Cross-national assessments alignment with the global framework for Reading and MPL analysis

This paper aims at answering the following questions:

1. How do the regional and international assessments of reading align with UNESCO Global Framework for Reading?
2. How can the alignment be improved?
3. How do the minimum proficiency levels from different regional and international assessments of reading relate to each other?
4. How can the comparability be improved?

In order to answer the previous questions an analysis of regional and international assessments of reading was carried out. Regarding the first question, the paper will show their alignment to the Global Framework at the three educational levels considered in the 4.1.1 indicator of SDG 4, which are Grades 2 & 3 (4.1.1a), the end of Primary education (4.1.1b) and at the end of Low Secondary education (4.1.1c).

Firstly, the cross-national assessments analyzed will be briefly described. Secondly, their alignment to the constructs from the Global Framework for Reading will be portrayed. Thirdly, the assessments will be compared in reference to their minimum proficiency levels, considering the possible overlapping between assessments designed for different educational levels. Finally, some recommendations for improving comparability will be presented.

Characteristics of the regional international assessments

Table 1 shows the cross-national assessments considered for this paper. Most of these tests are designed to evaluate formal learning, as is the case of reading. However, both ASER and UNICEF MICS 6 are broader questionnaires that aim at obtaining other development indicators at a personal, family, and environmental level, which include a section on reading that is the one considered in this analysis.

Table 1. Characteristics of the assessments

Name	Abbreviation	Grade/Age	Corresponding SDG 4 indicator	Minimum proficiency level	Observations
Annual Status of Education Report	ASER	6 to 14 year-olds	4.1.1.a;	Standard 2 (story)	Part of a household questionnaire in which the assessment is individual.
UNICEF Multiple Indicator Cluster Service	UNICEF MICS 6	5 to 17 year-olds	4.1.1.a;	Foundational Reading Skills	Part of a household questionnaire in which the assessment is individual.

Name	Abbreviation	Grade/Age	Corresponding SDG 4 indicator	Minimum proficiency level	Observations
UWEZO Annual Learning Assessment	UWEZO	6 to 16 year-olds	4.1.1.a;	Standard 2	Part of a household questionnaire in which the assessment is individual.
Early Grade Reading Assessment	EGRA	Grades 1 to 3.	4.1.1.a	Not specified	Individual assessment
Third Regional Comparative and Exploratory Study	TERCE	Grades 3 & 6	4.1.1.a; 4.1.1.b	Level 2	School-based assessment
Pacific Islands Literacy and Numeracy Assessment	PILNA	Grades 4 & 6	4.1.1.b	Level 4 (grade 4) and Level 5 (grade 6).	School-based assessment
Progress in International Reading Literacy Study	PIRLS	Grade 4	4.1.1.b	Low international Benchmark (second level)	School-based assessment
The Analysis Programme of the CONFEMEN Education Systems	PASEC	Grades 2 & 6	4.1.1.a; 4.1.1.b	Level 3	School-based assessment. Partly individual assessment
Southern and Eastern Africa Consortium for Monitoring Educational Quality	SACMEQ	Grade 6	4.1.1.b	Level 3	School-based assessment
Programme for International Student Assessment	PISA and PISA-D	15 year-olds	4.1.1.c	Level 2	School-based assessment

Regional and international assessments' alignment with the Global Framework for Reading

In order to answer the first question an analysis based on the performance level descriptors (PLDs) used by each assessment was carried out. The domains, sub domains and constructs belonging to the Global Framework for Reading considered by each assessment are shown in Table 2.

Table 2. Regional and international assessments' alignment with the Global Framework for Reading.

	READING COMPETENCY									LINGUISTIC COMPETENCY								METALINGUISTIC COMPETENCY			
	DECODING			READING COMPREHENSION						LISTENING			SPEAKING			VOCABULARY		PHONOLOGICAL AWARENESS			
ASSESSMENTS	1.1.1	1.1.2	1.1.3	1.2.1	1.2.2	1.2.3	1.2.4	1.2.5	1.2.6	2.1.1	2.1.2	2.1.3	2.2.1	2.2.2	2.2.3	2.3.1	2.3.2	3.3.1	3.3.2	3.3.3	3.3.4
ASER	X	X	X	X	X																
UNICEF MICS 6		X		X	X	X															
UWEZO	X	X	X	X	X																
EGRA	X	X	X	X	X	X				X	X						X	X	X		X
PASEC (Gr. 2)	X	X	X	X	X					X			X	X		X	X				X
TERCE (Gr.3)				X	X	X										X	X				
PIRLS (Gr. 4)			X	X	X	X	X														
PILNA (Gr. 4 y 6)				X	X	X	X														
PASEC (Gr. 6)		X	X	X	X	X	X														
TERCE (Gr. 6)				X	X	X	X														
SACMEQ (Gr. 6)				X	X	X	X										X				
PISA (Gr. 9)				X	X	X	X									X					

Note: In the Reading Competency the Decoding sub domain includes the following constructs 1.1.1. Alphabetic Principle; 1.1.2. Precision; 1.1.3. Fluency. In the Reading Comprehension sub domain are included: 2.1.1. Identify; 2.1.2. Retrieve; 2.1.3. Interpret; 2.1.4. Reflect; 2.1.4. Metacognition and 1.2.6. Motivation and Disposition. In the case of the Linguistic Competency, this includes the Listening sub domain which is constituted by 2.1.1 Retrieve; 2.1.2. Interpret and 2.1.3. Reflect. In the case of the Speaking sub domain, the constructs are 2.2.1. Form; 2.2.2. Content; 2.2.3. Use. The last sub domain from this competency is Vocabulary, being the constructs within 2.3.1. Acquire new words and 2.3.2. Recognize. Finally the Metalinguistic Competency has only one sub domain, phonological awareness, that includes the following constructs: 3.1.1. Distinguish; 3.1.2. Blend; 3.1.3. Generate words from and 3.1.4. Segment.

A common characteristic to all of the assessments analyzed is that independently of the age or grade they are designed for, these consider at least two of the constructs corresponding to the Reading Comprehension sub domain. This is of great relevance, as it is the only sub domain that is included in all of the assessments.

Moreover, it can be observed that the assessments designed to be applied in lower grades cover the Decoding sub domain, while the assessments designed for third grade onwards do not consider those constructs, except for PASEC in grade 6, which does include decoding in its assessment. This seems coherent with the developmental characteristics of reading acquisition. While the alphabetic principle is acquired in the first grade, in the second grade sufficiency levels are achieved regarding the precision construct. Finally, between third and fourth grade, depending on the characteristics of the language, sufficiency levels of fluency are attained.

Furthermore, regarding the Linguistic Competency, only five out of the twelve assessments analyzed include at least one of the constructs that correspond to this competency. PASEC (Grade 2) and EGRA stand out as the ones that more thoroughly evaluate this area. This is not surprising as reading acquisition implies the previous development of sufficiency levels both in the linguistic and metalinguistic competencies.

Finally, in reference to the Metalinguistic Competency, only EGRA and PASEC (grade 2) include the constructs belonging to this domain. It is important to consider that both of these assessments are either completely (EGRA) or partially (PASEC) applied individually, which allows to perform metaphonological tasks that would be almost impossible to conduct as a group. Moreover, both assessments present the evaluation of reading readiness as one of its aims; therefore it seems logical that phonological awareness tasks are included, as these are considered as pre-reading skills.

Comparison of minimum proficiency levels set by cross national assessments

In this section, in order to answer the third question a comparison between the different regional and international assessments is performed by looking for possible overlapping between assessments that are designed for different educational levels. This comparison is based on the minimum proficiency level (MPL) set by each assessment.

In this regard, even though most of the assessments state the specific grade or grades in which these should be applied, in some cases there seems to be incongruence between the performances expected by different assessments. Special cases are the ones of ASER, UNICEF MICS 6 and UWEZO, which have a broad age range of application.

A relevant aspect to take into consideration is that the MPL expected for TERCE in third grade (Level 2) is more demanding than all possible performance levels for PASEC (second grade). Moreover, this MPL is also more difficult than the first 5 levels of performance in SACMEQ (Grade 6) and PILNA (Grades 4 & 6). If we analyze the MPLs set for each assessment in table 1, it can be observed that TERCE's minimum proficiency for third grade is more difficult than what SACMEQ's and PILNA's for grade six. Moreover, all of the performance levels considered by ASER, UNICEF MICS 6 and UWEZO appear to be easier than the TERCE's lowest level for third grade. This is surprising considering that these assessments include students up to 14, 17 and 16 years old respectively, which means that the minimum proficiency expected for third graders by TERCE is higher than what is expected at the end of low secondary by these three assessments.

A similar situation is found when analyzing TERCE's MPL for Grade 6. When comparing Level 2 of TERCE, which is its MPL, with the other assessments designed for the same grade, it can be observed that TERCE's Level 2 performance descriptor is more difficult that all performance descriptors for PASEC, SACMEQ and PILNA. Furthermore, TERCE's Level 2 is also more difficult that PISA's Level 4, which is surprising considering that PISA sets its minimum proficiency in Level 2, and is designed for 15 year-olds. Once more this shows incongruence between the expectations that the different regional and international assessments have for students. Even though, in the case of TERCE the assessment is performed in Spanish, which is a transparent language, which could facilitate reading acquisition, the difference in language characteristics does not seem to be big enough to explain the expectation gap among assessments.

Similarly, but to a lesser extent, the first level of performance considered in PIRLS 2011 for grade 4 seems to be more difficult than the first two levels for PASEC (grade 6) and the first three levels for SACMEQ, also designed for grade 6. However, if we consider the first level of performance from PIRLS 2016 these differences sharpen, being more demanding that the first five levels from both PILNA and SACMEQ.

Finally, when analyzing PISA's MPL (Level 2) it can be observed that this level is slightly more difficult than the MPLs set by PASEC for grade 6, by PIRLS 2016 for grade 4, and by TERCE for grade 3. Moreover, as stated above it is easier than any of the performance level descriptors stated by TERCE for grade 6.

Conclusions and recommendations

In conclusion, if we go back to the first question that guided this paper it can be said that in general terms the regional and international assessments of reading align with UNESCO Global Framework for Reading. As expected given the characteristics of the assessments, this alignment is greater with the reading comprehension subdomain and to a lesser extent with the decoding subdomain. Few assessments explicitly evaluate the linguistic and metalinguistic competencies.

This seems coherent as some competencies are considered to have been previously acquired by students. However, in order to increase the alignment among the different regional and international assessments and between these and the Global Framework for Reading an option would be for the assessments to explicitly state in their framework which processes, abilities and skills are they assuming that students have already achieved when being evaluated.

Regarding the third question, even though most assessments are designed to be used with students in a specific grade or that are a specific age, the proficiency expected in the different assessments for the three educational levels is not necessarily congruent. Having found cases in which the expectations at the end of primary school are lower than those for grades 2 & 3, as well as cases in which the demands at the end of primary school are greater than those of the end of lower secondary.

This lack of congruence shows the need for the use of reference frameworks that are common to all of the assessments allowing for the international comparison of results.